CONTEXTUAL ONLINE ROUTING IN CONGESTED NETWORKS WITH BACKGROUND FLOW

Anonymous authors
Paper under double-blind review

ABSTRACT

We study contextual online routing in congested networks. In each round, a platform observes context and trip requests and routes divisible flow before observing the background flow of outside users. Edge costs are affine in total load, and noisy cost feedback comes only from used edges, with precision increasing with routed flow. Routing thus determines both congestion and where and how precisely costs are learned. Standard ellipsoidal optimism makes routing nonconvex. We propose RECTCOR (Rectangularized Contextual Optimistic Routing), which retains a convex quadratic routing problem by using the lower corner of the bounding rectangle of each projected confidence ellipse. Its statistical price is governed by the intercept–slope correlation of this ellipse and can cause linear regret when observed loads concentrate. Residual variation in background flow prevents this degeneracy uniformly over time and for any routing rule, yielding a high-probability $\widetilde{O}(\sqrt{T})$ regret bound without forced exploration. Background-flow variation is thus not only a source of uncertainty but also of excitation. In experiments, greater background-flow variation closes the regret gap between RECTCOR and ellipsoidal optimism, after which RECTCOR outperforms it at roughly one-tenth the computation; it also attains sublinear regret on the Sioux Falls network.

1 INTRODUCTION

Network flow must often be allocated before all congestion-inducing flow is observed. For example, a navigation platform routes a batch of trips after observing context, such as time of day, weather, and events, while sharing the network with outside drivers. We call their volume the background flow: it is unknown when routes are assigned and becomes available only afterward, for example through aggregated loop-detector counts (Chen et al., 2001). The platform's own flow is large enough to affect congestion, and cost feedback comes only from the edges it uses, with a precision that grows with the routed flow. Routing thus determines not only congestion but also where and how precisely the unknown edge costs are learned.

This setting creates both statistical and computational difficulties. Statistically, background flow is revealed only after execution, so routing relies on its prediction while edge costs are learned at realized loads. This resembles bandits with post-decision contextual information (Kirschner & Krause, 2019; Wang et al., 2023), except that here the post-decision variable combines with the chosen flow to determine the load at which the unknown cost is observed. Computationally, routing with known affine costs is a convex quadratic multi-commodity flow problem, but optimism over ellipsoidal confidence sets (Abbasi-Yadkori et al., 2011) can destroy this convexity by selecting different intercept–slope pairs at different loads.

Rectangular confidence sets have been used to retain a tractable routing problem for context-free affine congestion costs (Gollapudi et al., 2023), in a model with neither context nor background flow. With contextual costs, the confidence set must first be projected onto the current intercept–slope pair, and rectangularizing this projected set discards the dependence between the two coefficients. The resulting statistical price depends on the diversity of observed loads.

We propose RECTCOR (Rectangularized Contextual Optimistic Routing), which optimizes at the lower corner of the bounding rectangle of each projected cost confidence set. This keeps every routing decision a convex quadratic program while making the statistical price of tractability explicit.

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

Our analysis shows that fluctuations in background flow both hinder and help learning: they make decision-time congestion harder to predict, but also diversify the loads at which edge costs are observed, which keeps the statistical price of rectangularization bounded. Background-flow variation is thus not only a source of uncertainty to be predicted but also a source of excitation for learning costs. Unlike context diversity, which makes greedy policies efficient in linear contextual bandits (Kannan et al., 2018; Bastani et al., 2021), this excitation arises after the decision, in the loads at which costs are observed, and it bounds the price of the rectangle rather than replacing optimism.

Our main contributions are as follows:

- **Problem formulation.** In Section 2, we formulate contextual online routing on a congested network. The platform routes divisible flow before background flow is revealed on every edge, and observes noisy costs only on used edges, with a precision that increases with the routed flow.
- **Tractable optimistic algorithm.** In Section 3, we develop RECTCOR, which combines full-feedback prediction of background flow with precision-weighted regression of edge costs at the realized loads. It maps each cost confidence ellipsoid exactly onto the current intercept–slope pair and takes the lower corner of its bounding rectangle, a single optimistic cost that is valid for all loads. Every routing decision is then a convex quadratic program of polynomial size.
- **The price of rectangularization.** In Section 4.2, we quantify what is lost by replacing the projected confidence ellipse with a rectangle. In the load directions relevant to routing, the cost gap is inflated by the factor $\sqrt{2/(1+\rho_{t,e})}$, where $\rho_{t,e}$ is the intercept–slope correlation of the projected confidence set (Lemma 1). The untruncated lower corner is the greatest convex minorant of the ellipsoidal lower bound, so this factor is the price of convexity itself.
- **Background flow as a source of excitation.** In Section 4.3, we show that nondegenerate residual variation of background flow keeps $1+\rho_{t,e}$ bounded away from zero (Proposition 1). This holds with high probability, uniformly in time, and for any non-anticipating routing rule. Together with the price of rectangularization, this yields a high-probability $\widetilde{\mathcal{O}}(\sqrt{T})$ regret bound without forced exploration (Theorem 1).
- **Empirical evaluation.** In Section 5, we empirically show that background flow acts as a source of excitation. On a small network, the regret of RECTCOR decreases as background flow varies more, most sharply when it is the only source of load variation. Its gap to ellipsoidal optimism closes as this variation grows, and RECTCOR attains lower regret at the largest variation, whereas ellipsoidal optimism requires solving a nonconvex routing problem and is 8–16 times slower per round. On the Sioux Falls network, RECTCOR attains sublinear regret and outperforms ε_t-greedy.

1.1 RELATED WORK

Online routing and combinatorial semi-bandits. The online shortest-path problem provides an early formulation of routing with partial edge-level feedback (György et al., 2007). More generally, combinatorial semi-bandits exploit component-level observations from structured actions rather than treating each feasible action as an independent arm (Gai et al., 2012; Chen et al., 2013; Kveton et al., 2015). Linear generalization has since been incorporated into large-scale and contextual combinatorial semi-bandits (Wen et al., 2015; Takemura et al., 2021), with more recent work considering general function approximation (Qin & Zhang, 2026). Our problem shares their component-level feedback, but an edge has no fixed unknown mean: the routed flow sets the load at which its cost is evaluated, so the action determines both whether an edge is observed and the covariate of the observation. Decisions are also continuous multi-commodity flows, not discrete subsets or paths.

Divisible allocation and allocation-dependent information. A related literature studies bandit problems in which a divisible resource is distributed among several alternatives, and the allocation itself affects the available statistical information. Lattimore et al. (2014) study resource allocation with semi-bandit feedback, and subsequent work considers multiple resources and linear structure (Lattimore et al., 2015). Li & Chen (2023) study divisible-resource allocation when rewards are unknown and random, while heteroskedastic bandit models account for observations whose noise levels vary between decisions (Kirschner & Krause, 2018). These models relate to our flow-dependent precision, but here flow also changes the unknown mean cost through congestion and moves the load at which the cost is sampled, which background flow shifts further after the allocation.

Learning and decision-making under congestion. Another complementary line of work studies repeated interaction in congestion and routing games. Blum et al. (2006) analyze regret-minimizing dynamics and their relation to equilibrium in routing games, while Cui et al. (2022) study learning in congestion games under bandit feedback. Awasthi et al. (2022) consider online routing when rewards depend on recent congestion history and formulate guarantees through policy regret. We instead consider a centralized platform controlling divisible requests and compare its decisions with a stagewise, context-aware non-clairvoyant oracle. Our goal is not to reach an equilibrium or to compete with a history-dependent policy, but to quantify the cost of learning contextual congestion functions and predicting exogenous flow while repeatedly solving a network allocation problem.

2 PRELIMINARIES

2.1 PROBLEM FORMULATION

We consider a flow network represented by a directed multigraph $G = (V, E)$, where V is the set of nodes and E is the set of directed edges. For a rigorous definition of a directed multigraph, see Appendix A.1. Figure 1 illustrates the setting.

Let $\mathcal{I}$ be the set of feasible origin–destination (OD) pairs $i = (o_i, d_i) \in V \times V$, where feasibility means that at least one directed path connects o_i to d_i. Let $\mathcal{P}_i$ denote the set of simple paths for OD pair i, and let $\mathcal{P} = \bigcup_{i\in\mathcal{I}} \mathcal{P}_i$. We restrict attention to simple paths, excluding paths with cycles.

In each round $t = 1, \dots, T$, the learner observes a context $X_t \in \mathcal{X}$ and a request collection $R_t \in \mathbb{R}_+^{\mathcal{I}}$, where $\mathcal{X}$ is compact. The total controlled demand $M_t = \sum_{i\in\mathcal{I}} R_{t,i}$ is positive and bounded. The learner then selects a divisible path flow $f_t \in \mathcal{F}(R_t)$, where $\mathcal{F}(r) = \left\{ f \geq 0 : \sum_{p\in\mathcal{P}_i} f_p = r_i \quad \forall i \in \mathcal{I} \right\}$. The feasible set $\mathcal{F}(r)$ is a nonempty compact polytope. For a path flow f, the controlled load on edge e is $n_e(f) = \sum_{p:e\in p} f_p$, and we write $n_{t,e} = n_e(f_t)$.

In addition to the controlled flow, the network carries nonnegative background flow $U_t = \{U_{t,e}\}_{e\in E}$. The current U_t is not available when f_t is chosen and is revealed on all edges only after execution. We use a linear feature model for the conditional mean of background flow. Let $\phi_u : \mathcal{X} \to \mathbb{R}^{d_u}$ be a known feature map, and define

$$U_{t,e} = \mu_e(X_t) + \varepsilon_{t,e} \geq 0, \qquad \mu_e(x) = \phi_u(x)^\top \gamma_e \geq 0, \tag{1}$$

where $\gamma_e \in \mathbb{R}^{d_u}$ is unknown. The residual $\varepsilon_{t,e}$ is centered conditional on the information available before execution, so that $\mu_e(X_t)$ is the conditional mean of $U_{t,e}$.

Each edge has a context-dependent unit-cost function that is affine in its total load. Let $\phi_c : \mathcal{X} \to \mathbb{R}^{d_c}$ be a known feature map, and suppose

$$c_e(x, z) = a_e(x) + b_e(x)z = \phi_c(x)^\top(\beta_{e,1} + z\,\beta_{e,2}), \qquad z \geq 0,$$

where $\beta_{e,1}, \beta_{e,2} \in \mathbb{R}^{d_c}$ are unknown. We assume that the base cost $a_e(x)$ and the slope $b_e(x)$ are nonnegative for all $x \in \mathcal{X}$ and $e \in E$, and that $\phi_c(x)$ contains a constant component (thus nonzero).

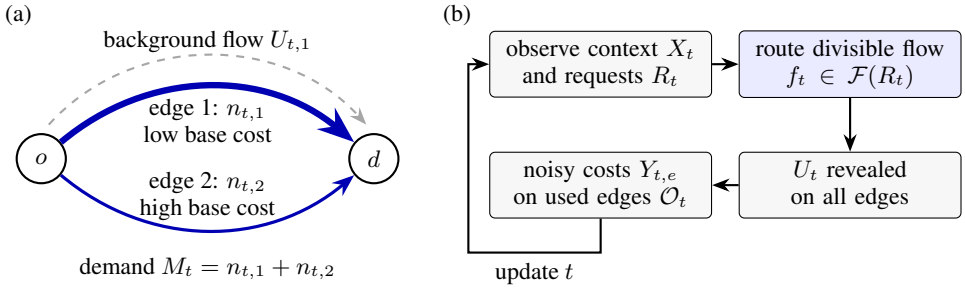


Figure 1: (a) The platform splits its demand M_t (blue, width proportional to load) between two parallel edges; the cheaper edge 1 attracts flow until its congestion, including background flow (gray, dashed), offsets its advantage. (b) One round: routing precedes the background flow, and costs are observed only on edges that carry controlled flow.

After executing the decision f_t, define the total load realized on edge e by $Z_{t,e} = n_{t,e} + U_{t,e}$. Cost feedback is semi-bandit; the learner observes a noisy unit cost only on the set of used edges $\mathcal{O}_t = \{e \in E : n_{t,e} > 0\}$. For each $e \in \mathcal{O}_t$,

$$Y_{t,e} = c_e(X_t, Z_{t,e}) + \xi_{t,e}. \tag{2}$$

The learner fixes nonnegative calibration constants τ_0, τ_1, not both zero, and uses the noise-scale profile $\bar{v}(n) = \tau_0^2 + \tau_1^2/n$. The algorithm uses the corresponding precision $\omega(n) = 1/\bar{v}(n)$ when fitting the cost model. Write $a_{t,e} = a_e(X_t)$, $b_{t,e} = b_e(X_t)$, and $\mu_{t,e} = \mu_e(X_t)$.

Why congestion makes routing nontrivial. Without congestion, that is, with $b_{t,e} = 0$ on every edge, each unit cost is fixed, and the best decision sends every commodity along its cheapest path. In a congested network, however, unit costs increase with the load, so sending all flow along the path with the lowest base cost congests it, and the demand must be split. For a commodity with two parallel edges (Figure 1), the split of the controlled demand $M_t = n_{t,1} + n_{t,2}$ that minimizes the expected cost of the controlled flow (Section 2.2) equalizes the marginal costs of the two edges,

$$a_{t,1} + b_{t,1}(2n_{t,1} + \mu_{t,1}) = a_{t,2} + b_{t,2}(2n_{t,2} + \mu_{t,2}),$$

whenever both are used. An edge with a low base cost thus attracts flow only until its congestion offsets this advantage, and heavy background flow on an edge pushes controlled flow away from it. The learner must strike this balance without knowing the costs or the mean background flow, and each split also determines the loads at which the costs are observed.

2.2 Objective

The platform evaluates only the cost incurred by its controlled users. For a feasible path flow f, context x, and realized background flow u, define $C(f; x, u) = \sum_{e \in E} n_e(f) c_e\big(x, n_e(f) + u_e\big)$. Note that background flow affects congestion but is not itself weighted in the platform's objective. Define $q_{t,e}(n) = a_{t,e} + b_{t,e}(n + \mu_{t,e})$, $C_t(f) = \sum_{e \in E} n_e(f) q_{t,e}\big(n_e(f)\big)$, and $L_t(f) = C_t(f)/M_t$. Here, $C_t(f)$ is the conditional expected total cost of the controlled flow given the information available before the decision, and $L_t(f)$ is the average cost per unit of controlled demand.

We compare the learner with the optimal benchmark decision $f_t^\star \in \arg\min_{f \in \mathcal{F}(R_t)} L_t(f)$. Define the cumulative regret by

$$\text{Regret}_T = \sum_{t=1}^{T} \left[L_t(f_t) - L_t(f_t^\star) \right]. \tag{3}$$

The benchmark knows the model parameters but, like the learner, observes only (X_t, R_t) before deciding, since U_t is revealed only after the routing decision.

3 Algorithm

RectCOR (Algorithm 1) predicts background flow by full-feedback ridge regression, learns edge costs by precision-weighted ridge regression at the realized loads, and routes with an optimistic affine cost taken from the lower corner of a rectangle around the current intercept and slope. Appendix A.2 gives the incremental updates and further details.

Two regressions with distinct feedback. Background flow is observed on every edge, so all edges share the context design:

$$V_t^u = \lambda_u I + \sum_{s<t} \phi_u(X_s)\phi_u(X_s)^\top, \qquad h_{t,e}^u = \sum_{s<t} \phi_u(X_s) U_{s,e}, \qquad \widehat{\gamma}_{t,e} = (V_t^u)^{-1} h_{t,e}^u, \tag{4}$$

and RectCOR uses $\widetilde{\mu}_{t,e} = [\widehat{\mu}_{t,e}]_+$ with $\widehat{\mu}_{t,e} = \phi_u(X_t)^\top \widehat{\gamma}_{t,e}$. Edge costs are observed only on used edges, with precision $\omega(n) = (\tau_0^2 + \tau_1^2/n)^{-1}$. Write $\psi(x, z) = (\phi_c(x), z\phi_c(x))^\top$ and $\beta_e = (\beta_{e,1}, \beta_{e,2})^\top$, so that $c_e(x, z) = \psi(x, z)^\top \beta_e$, and for $e \in \mathcal{O}_s$ let $\chi_{s,e} = \sqrt{\omega(n_{s,e})}\, \psi(X_s, Z_{s,e})$ and $y_{s,e}^c = \sqrt{\omega(n_{s,e})}\, Y_{s,e}$. The precision-weighted ridge estimator is

$$V_{t,e}^c = \lambda_c I + \sum_{s<t:\, e \in \mathcal{O}_s} \chi_{s,e}\chi_{s,e}^\top, \qquad h_{t,e}^c = \sum_{s<t:\, e \in \mathcal{O}_s} \chi_{s,e} y_{s,e}^c, \qquad \widehat{\beta}_{t,e} = (V_{t,e}^c)^{-1} h_{t,e}^c. \tag{5}$$

It uses the realized load $Z_{s,e} = n_{s,e} + U_{s,e}$, available once background flow is revealed.

Algorithm 1 RECTCOR: Rectangularized Contextual Optimistic Routing

Require: regularization $\lambda_u, \lambda_c > 0$; calibration constants τ_0, τ_1; exploration radii $\eta_{t,e}$.
1: Initialize $V_1^u = \lambda_u I_{d_u}$ and, for all $e \in E$, $h_{1,e}^u = 0$, $V_{1,e}^c = \lambda_c I_{2d_c}$, and $h_{1,e}^c = 0$.
2: **for** $t = 1, \dots, T$ **do**
3: Observe context X_t and requests R_t.
4: **(Background flow)** For all e, predict $\widetilde{\mu}_{t,e} = [\phi_u(X_t)^\top \widehat{\gamma}_{t,e}]_+$ from the ridge estimate Eq. (4).
5: **(Optimistic cost)** For all e, project the cost confidence ellipsoid onto the current intercept and slope, and take the lower corner $(\underline{a}_{t,e}, \underline{b}_{t,e})$ of its bounding rectangle, as in Eq. (7).
6: **(Routing)** Choose f_t by solving the convex quadratic program Eq. (8) with these optimistic costs and the predicted background flow.
7: Execute f_t; observe $U_{t,e}$ for all e and $Y_{t,e}$ for $e \in \mathcal{O}_t$.
8: **(Update)** Add the new contexts and the precision-weighted cost observations at the realized loads $Z_{t,e} = n_{t,e} + U_{t,e}$ to the ridge statistics, as in Eq. (12).
9: **end for**

Current-context optimistic reduction. Routing at context X_t depends on β_e only through the current intercept and slope $\zeta_{t,e} = A_t \beta_e = (a_{t,e}, b_{t,e})^\top$, where $A_t = \left(\begin{smallmatrix} \phi_c(X_t)^\top & 0 \\ 0 & \phi_c(X_t)^\top \end{smallmatrix}\right)$. Let $\widehat{\zeta}_{t,e} = A_t \widehat{\beta}_{t,e} = (\widehat{a}_{t,e}, \widehat{b}_{t,e})^\top$ and $G_{t,e} = A_t (V_{t,e}^c)^{-1} A_t^\top$. For a radius $\eta_{t,e} \geq 0$, the image of the ellipsoid $\mathcal{C}_{t,e}^\beta = \{\beta : \|\beta - \widehat{\beta}_{t,e}\|_{V_{t,e}^c} \leq \eta_{t,e}\}$ under A_t is exactly

$$\mathcal{C}_{t,e}^\zeta = A_t \mathcal{C}_{t,e}^\beta = \left\{\zeta : \|\zeta - \widehat{\zeta}_{t,e}\|_{G_{t,e}^{-1}} \leq \eta_{t,e}\right\}, \tag{6}$$

so the projection adds no confidence inflation. RECTCOR takes the nonnegative lower corner of the bounding rectangle of this ellipse,

$$\underline{a}_{t,e} = \left[\widehat{a}_{t,e} - \eta_{t,e}\sqrt{(G_{t,e})_{11}}\right]_+, \qquad \underline{b}_{t,e} = \left[\widehat{b}_{t,e} - \eta_{t,e}\sqrt{(G_{t,e})_{22}}\right]_+. \tag{7}$$

If $\beta_e \in \mathcal{C}_{t,e}^\beta$, then $0 \leq \underline{a}_{t,e} \leq a_{t,e}$ and $0 \leq \underline{b}_{t,e} \leq b_{t,e}$, so the single affine cost $\underline{c}_{t,e}(z) = \underline{a}_{t,e} + \underline{b}_{t,e} z$ is optimistic for every load $z \geq 0$. Section 4 calibrates $\eta_{t,e}$ so that this holds with high probability.

Optimistic convex routing. RECTCOR chooses

$$f_t \in \operatorname*{arg\,min}_{f \in \mathcal{F}(R_t)} \sum_{e \in E} n_e(f) \left[\underline{a}_{t,e} + \underline{b}_{t,e}\big(n_e(f) + \widetilde{\mu}_{t,e}\big)\right]. \tag{8}$$

Since $\underline{a}_{t,e}, \underline{b}_{t,e}, \widetilde{\mu}_{t,e} \geq 0$, this is a convex quadratic program in every round. Appendix B gives a polynomial-size formulation that never enumerates paths and a Frank–Wolfe implementation based on shortest paths (Jaggi, 2013). Minimizing the cost pointwise over the projected ellipse would lose convexity, because different loads could then select different intercept–slope pairs.

4 REGRET ANALYSIS

We bound the regret of RECTCOR against the non-clairvoyant benchmark of Section 2.2. Let $\mathcal{H}_{t-1}$ contain the contexts, decisions, and feedback of the previous rounds, and let $\mathcal{D}_t^0 = \sigma(\mathcal{H}_{t-1}, X_t, R_t)$, $\mathcal{D}_t = \sigma(\mathcal{D}_t^0, f_t)$, and $\mathcal{G}_t = \sigma(\mathcal{D}_t, U_t)$ denote the information before the decision, after the decision, and after observing the background flow. The decision f_t is $\mathcal{D}_t^0$-measurable.

4.1 ASSUMPTIONS AND CONFIDENCE SETS

Assumption 1 (Boundedness) *There exist finite constants $B_\phi, B_\theta > 0$ such that, for every $x \in \mathcal{X}$ and $e \in E$, $\|\phi_u(x)\|_2, \|\phi_c(x)\|_2 \leq B_\phi$ and $\|\gamma_e\|_2, \|\beta_e\|_2 \leq B_\theta$.*

Let $B_0 = B_\phi B_\theta$, so that $a_e(x), b_e(x), \mu_e(x) \in [0, B_0]$ for all x and e, and denote the bounds on the controlled demand by $0 < \underline{m} \leq M_t \leq \overline{m} < \infty$.

Assumption 2 (Sub-Gaussian noise of background flow) *For some finite $\sigma \geq 0$, the residual in Eq. (1) satisfies, for every $s \in \mathbb{R}$,*

$$\mathbb{E}[\varepsilon_{t,e} \mid \mathcal{D}_t] = 0, \qquad \mathbb{E}[\exp(s\varepsilon_{t,e}) \mid \mathcal{D}_t] \leq \exp(s^2\sigma^2/2).$$

In particular, $\mathbb{E}[U_{t,e} \mid \mathcal{D}_t] = \mu_e(X_t)$: the routing decision does not affect the conditional mean of the background flow.

Assumption 3 (Flow-dependent, sub-Gaussian cost noise) *For every used edge $e \in \mathcal{O}_t$, the noise in Eq. (2) satisfies, for every $s \in \mathbb{R}$,*

$$\mathbb{E}[\xi_{t,e} \mid \mathcal{G}_t] = 0, \qquad \mathbb{E}[\exp(s\xi_{t,e}) \mid \mathcal{G}_t] \leq \exp\left\{ \frac{s^2}{2} \left(\tau_0^2 + \frac{\tau_1^2}{n_{t,e}} \right) \right\}.$$

Assumptions 1 and 2 are standard in linear bandits (Abbasi-Yadkori et al., 2011): bounded features and parameters, and conditionally sub-Gaussian noise. Assumption 3 is less standard, since its variance proxy decreases with the routed flow. The reported cost is the average travel time of the $n_{t,e}$ routed users: noise shared by all of them contributes τ_0^2, while each user's own noise averages out to $\tau_1^2/n_{t,e}$. Heteroskedastic bandits (Kirschner & Krause, 2018) study similar action-dependent noise. The assumption conditions on $\mathcal{G}_t$, which already contains the background flow U_t, so the regression can use the realized covariate $\psi(X_t, Z_{t,e})$ even though $Z_{t,e}$ may be unbounded.

Confidence radii. Let $\Gamma_t^u = \log[\det(V_t^u)/\det(\lambda_u I)]$ and $\Gamma_{t,e}^c = \log[\det(V_{t,e}^c)/\det(\lambda_c I)]$. For $\delta \in (0,1)$, define

$$\alpha_t^u(\delta) = \sqrt{\lambda_u} B_\theta + \sigma\sqrt{\Gamma_t^u + 2\log(|E|/\delta)}, \qquad \alpha_{t,e}^c(\delta) = \sqrt{\lambda_c} B_\theta + \sqrt{\Gamma_{t,e}^c + 2\log(|E|/\delta)}. \quad (9)$$

Under Assumptions 1–3, the self-normalized bound of Abbasi-Yadkori et al. (2011) implies that, with probability at least $1-\delta$, $\beta_e \in \mathcal{C}_{t,e}^\beta$ for all $t \geq 1$ and $e \in E$ when $\eta_{t,e} = \alpha_{t,e}^c(\delta)$. The same holds for γ_e with radius $\alpha_t^u(\delta)$ (Appendix C).

4.2 The price of rectangularization

The lower corner bounds the intercept and slope separately, which discards their dependence. Define $\rho_{t,e} = (G_{t,e})_{12}/\sqrt{(G_{t,e})_{11}(G_{t,e})_{22}}$. Then, for every $z \geq 0$,

$$\sqrt{(G_{t,e})_{11}} + z\sqrt{(G_{t,e})_{22}} \leq \sqrt{\frac{2}{1+\rho_{t,e}}} \sqrt{(G_{t,e})_{11} + 2z(G_{t,e})_{12} + z^2(G_{t,e})_{22}}, \quad (10)$$

with equality at $z = \sqrt{(G_{t,e})_{11}/(G_{t,e})_{22}}$. The two sides are the widths of the rectangle and of the projected ellipse Eq. (6) in the direction $(1, z)$. Since routing uses only $z \geq 0$, the factor diverges as $\rho_{t,e} \to -1$ and approaches one as $\rho_{t,e} \to +1$. This yields a bound on the optimistic cost gap $g_{t,e}(z) = c_e(X_t, z) - \underline{c}_{t,e}(z)$.

Lemma 1 (Directional price of rectangularization) *Fix t and e, and suppose $\beta_e \in \mathcal{C}_{t,e}^\beta$. Then $0 \leq \underline{a}_{t,e} \leq a_{t,e}$ and $0 \leq \underline{b}_{t,e} \leq b_{t,e}$, so $g_{t,e}(z) \geq 0$ for all $z \geq 0$. If moreover $1 + \rho_{t,e} \geq \kappa$ for some $\kappa \in (0,1]$, then for every $z \geq 0$,*

$$g_{t,e}(z) \leq \min\left\{ K(\kappa)\, \eta_{t,e} \|\psi(X_t, z)\|_{(V_{t,e}^c)^{-1}},\ B_0(1+z) \right\}, \qquad K(\kappa) = 1 + \sqrt{2/\kappa}.$$

In $K(\kappa)$, the term 1 is the error of the projected center, which the ellipsoidal bound also incurs, and $\sqrt{2/\kappa}$ is the additional width of the rectangle. Neither optimism nor convexity requires $1 + \rho_{t,e}$ to be bounded below; this is needed only to control the regret.

The rectangle is not an arbitrary relaxation. The ellipsoidal lower bound $\ell_{t,e}(z) = \min_{\zeta \in \mathcal{C}_{t,e}^\zeta} (1, z)\zeta$ is concave in z, and the untruncated lower corner $\ell_{t,e}^\star(z) = (\widehat{a}_{t,e} - \eta_{t,e}\sqrt{(G_{t,e})_{11}}) + (\widehat{b}_{t,e} - \eta_{t,e}\sqrt{(G_{t,e})_{22}})z$ is its greatest convex minorant on $[0, \infty)$ (Lemma 13). Hence any optimistic cost

that is convex in the load and valid for the whole confidence set lies below the lower corner, and $\sqrt{2/(1+\rho_{t,e})}$ is the price of convexity itself. This parallels linear bandits, where replacing the ellipsoid by an ℓ_1 ball restores tractability at the cost of a $\sqrt{d}$ factor (Dani et al., 2008); here the price depends on $\rho_{t,e}$. Without a lower bound on $1+\rho_{t,e}$, the rectangle can fail: in the instance of Appendix C.6, repeated observations at a single load drive $\rho_{t,e} \to -1$, the lower corner stays at zero, and the regret grows linearly.

4.3 Compatibility supplied by background flow

Assumption 4 (Residual variation) *There exists a deterministic $v_* > 0$ such that, for every $t \geq 1$ and $e \in \mathcal{O}_t$,*

$$\mathbb{E}[(\varepsilon_{t,e})^2 \mid \mathcal{D}_t] = \mathrm{Var}(U_{t,e} \mid \mathcal{D}_t) \geq v_*.$$

The bound concerns variability that remains after context, history, demand, and the decision are fixed. It is needed only on used edges, where, since $U_{t,e} \geq 0$, it requires $\mu_e(X_t)$ to be bounded away from zero.

Proposition 1 (Time-uniform compatibility from residual variation) *Suppose Assumptions 1, 2, and 4 hold with $(\tau_0, \tau_1) \neq (0,0)$, and let $h = \omega(\overline{m})B_\phi^2$. For $\delta \in (0,1)$ and $\lambda_0 > 0$, set*

$$\lambda_c = \lambda_0 h \log \frac{4d_c|E|}{\delta}. \tag{11}$$

Then there exists a constant $\kappa_* \in (0,1)$, depending only on $\overline{m} + B_0$, σ, v_*, and λ_0 and given explicitly in Eq. (19), such that for any non-anticipating routing rule using the weighted cost design of Section 3,

$$\mathbb{P}\big\{1 + \rho_{t,e} \geq \kappa_* \text{ for all } t \geq 1,\ e \in E\big\} \geq 1 - \delta.$$

The proof (Appendix C.4) sandwiches the realized cost design between two multiples of the design with separately observed intercept and slope, so $G_{t,e}$ has a bounded condition number. For fixed λ_0 and $Z_0 = \overline{m} + B_0$, κ_* increases with v_*. When $v_* \propto \sigma^2$, κ_* grows like σ^2 for small σ and decays polynomially for large σ, since heavier tails also inflate the realized design, so larger variance need not always reduce regret. Residual variation enters the analysis only through this compatibility event: the explicit bound of Proposition 3 holds whenever $1+\rho_{t,e} \geq \kappa$ along the trajectory, whatever supplies it, such as the demand and context variation in our Sioux Falls experiment (Appendix D.3).

4.4 Regret bound

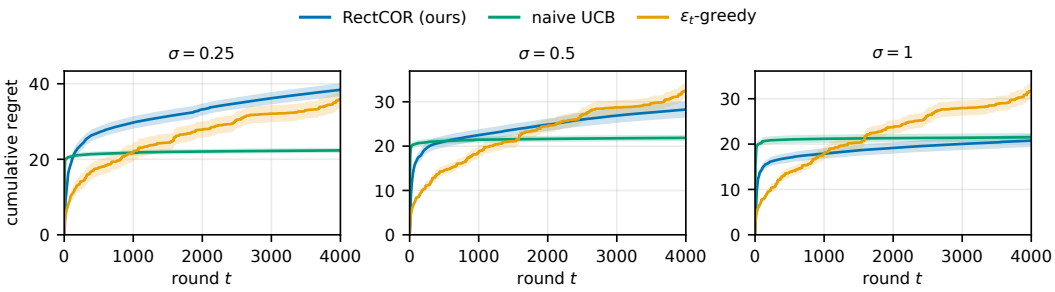
Let $H_G = \max_{p \in \mathcal{P}} |p|$ denote the largest number of edges on a simple path. We assume throughout that the routing program Eq. (8) is solved exactly; an inexact solver adds its accumulated optimization error to the bound (Remark 2).

Theorem 1 (Regret bound) *Suppose Assumptions 1–4 hold with $(\tau_0, \tau_1) \neq (0,0)$. For $\delta \in (0,1)$ and $\lambda_0 > 0$, run Algorithm 1 with $\lambda_c = \lambda_0 h \log(20 d_c|E|/\delta)$ and $\eta_{t,e} = \alpha^c_{t,e}(\delta/5)$. Then, with probability at least $1-\delta$,*

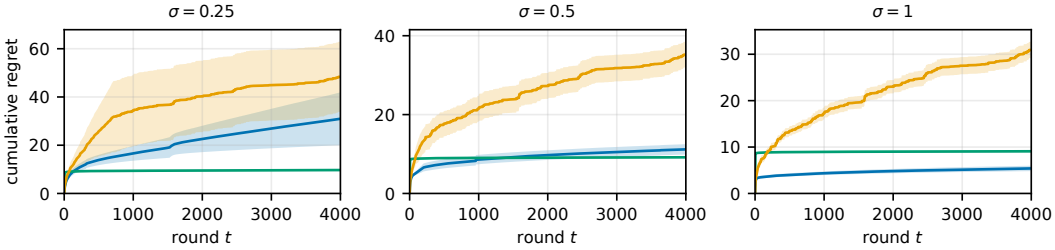
$$\textit{Regret}_T = \widetilde{\mathcal{O}}\Big(\big(\kappa_*^{-1/2} d_c \sqrt{H_G|E|} + H_G d_u\big)\sqrt{T} + d_c|E|\Big),$$

where κ_* is the constant of Proposition 1 and $\widetilde{\mathcal{O}}$ hides factors polylogarithmic in T, $|E|$, d_c, and $1/\delta$, as well as constants depending only on B_ϕ, B_θ, σ, τ_0, τ_1, $\underline{m}$, $\overline{m}$, λ_u, and λ_0.

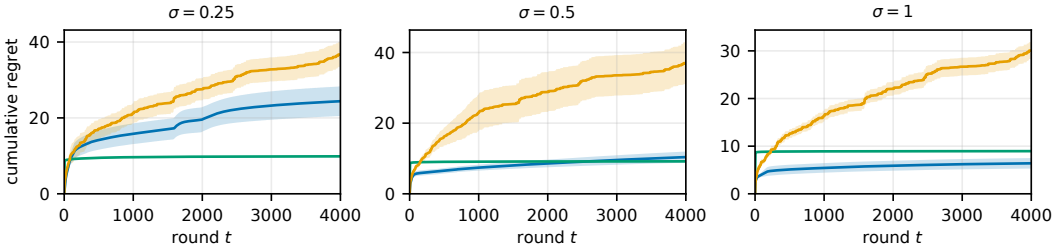
Interpretation. The three terms are the cost-learning error, the flow-prediction error (the price of deciding before background flow is revealed), and a lower-order burn-in term (Appendix C.1). Three features deserve emphasis. First, convexity costs only a constant factor: for ellipsoidal, nonconvex optimism, the factor $K(\kappa)$ of Lemma 1 would be 2, so rectangularization inflates the leading term by the factor $K(\kappa_*)/2 = O(\kappa_*^{-1/2})$ and leaves the $\sqrt{T}$ rate unchanged. This factor is set by the environment rather than by the algorithm, which needs neither forced exploration nor knowledge of v_*. Second, the bound depends on the network only through $|E|$ and the path length H_G, not through



(a) Varying context and demand.



(b) Fixed context, varying demand.



(c) Fixed context and demand.

Figure 2: Cumulative regret on the small network for background-flow residual scale $\sigma \in \{0.25, 0.5, 1\}$ (columns), $T = 4000$, mean $\pm$ one standard error over 20 seeds. ε_t-greedy uses $c = 0.25$, the best value in $\{0.25, 0.5, 1, 2\}$.

the number of paths, which can be exponential in $|E|$; nor does it require independence across edges. Third, the $\sqrt{T}$ dependence matches the optimal rate for stochastic linear bandits (Dani et al., 2008). Proposition 4 shows that RECTCOR needs some compatibility condition; whether every computationally efficient algorithm must pay a similar price remains open.

Proof sketch. Under optimism (Lemma 1), the one-step regret is at most the cost gap at the predicted loads $n_{t,e} + \mu_{t,e}$ plus a flow-prediction term, so no width at the benchmark's loads is needed, and the flow term sums to $\widetilde{\mathcal{O}}(H_G d_u \sqrt{T})$. The cost design, however, is updated at the realized loads $Z_{t,e} = n_{t,e} + \mu_{t,e} + \varepsilon_{t,e}$. Since $g_{t,e}$ is affine, $g_{t,e}(Z_{t,e}) - g_{t,e}(n_{t,e} + \mu_{t,e}) = (b_{t,e} - \underline{b}_{t,e})\varepsilon_{t,e}$, a martingale difference that needs no independence across edges. Lemma 1 with $\kappa = \kappa_*$ and the elliptical potential of the realized design bound the rest (Appendix C).

5 NUMERICAL EXPERIMENTS

We evaluate RECTCOR on a small network, where each source of load variation can be switched off, and on the Sioux Falls road network (LeBlanc et al., 1975; Transportation Networks for Research Core Team, 2016). The small network has 5 nodes, 7 edges, and two commodities sharing a bottleneck edge. Context and demand vary in Type 1 (Figure 2a), only demand varies in Type 2 (Figure 2b), and background flow is the only source of load variation in Type 3 (Figure 2c); each

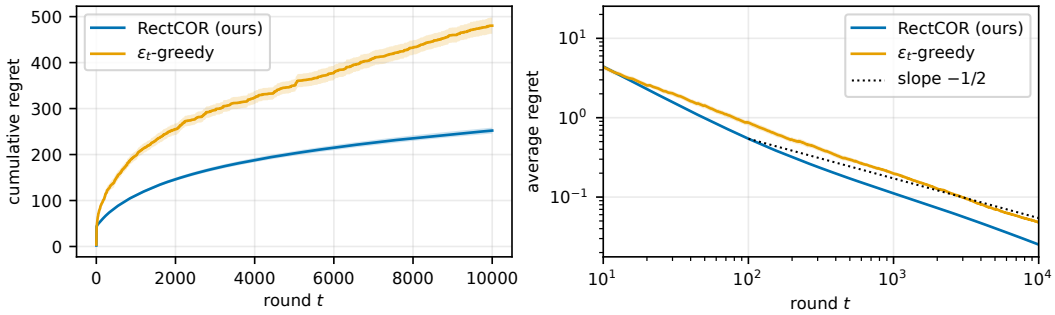


Figure 3: Sioux Falls network, $T = 10000$, mean $\pm$ one standard error over 20 seeds. Left: cumulative regret. Right: average regret Regret_t/t on log–log axes, with a reference line of slope $-1/2$.

type uses $\sigma \in \{0.25, 0.5, 1\}$. We compare against ε_t-greedy and naive UCB, which uses the ellipsoidal lower bound $\min_{\zeta \in C^{\zeta}_{t,e}} (1, z)\zeta$ instead of the rectangle. This bound is concave in the load, so the routing problem is nonconvex, and concave-cost network flow is NP-hard in general (Guisewite & Pardalos, 1990); we solve it globally on the small network. Appendix D gives all settings.

Results. Larger background-flow variation lowers the regret of RECTCOR in every type, most sharply when it is the only source of variation: in Type 3, the mean cumulative regret at $T = 4000$ falls from 24.4 to 6.4 as σ increases from 0.25 to 1 (Figure 2). Naive UCB does not rely on compatibility; its regret is nearly constant in σ. It is better at $\sigma = 0.25$, but the gap closes as σ grows, and at $\sigma = 1$, RECTCOR attains lower regret in all three types. This is the price of rectangularization shrinking with residual variation, as Lemma 1 and Proposition 1 predict. Naive UCB pays for its statistical efficiency in computation: its routing problem was nonconvex in 95–100% of rounds, and it took 2.97 ms per round against 0.30 ms for RECTCOR, even on a network with six paths. RECTCOR also beats ε_t-greedy in eight of the nine configurations (Table 2 in Appendix D). On the Sioux Falls network, with 24 nodes, 76 edges, and 10 OD pairs, RECTCOR attains sublinear regret, 252.1 ± 6.4 at $T = 10000$ against 480.3 ± 17.6 for ε_t-greedy (Figure 3).

6 CONCLUSION

We studied contextual online routing in congested networks, where a platform routes divisible flow before observing background flow and learns edge costs only where it routes. RECTCOR keeps every routing decision a convex quadratic program by taking a single optimistic corner of the exact intercept–slope projection of each confidence set. Its statistical price is the price of convexity itself, governed by the intercept–slope correlation of the projected set; it can cause linear regret when observed loads concentrate, but residual variation of background flow keeps it bounded uniformly in time for any routing rule, which yields an $\widetilde{O}(\sqrt{T})$ regret bound. Background flow thus acts not only as a nuisance to be predicted but also as a source of excitation. In experiments, RECTCOR matches or improves on the regret of nonconvex ellipsoidal optimism as background flow varies more, at about a tenth of its computation per round. Since the analysis uses background flow only through the compatibility event, other sources of compatibility, such as demand variation or planned exploration, yield the same guarantee. Natural extensions include nonlinear congestion costs, partially observed background flow, and background flow that responds to the platform's routing.

AI USE STATEMENT

In this work, we used generative AI tools for assistance in the development and verification of mathematical proofs and in the implementation of simulation experiments. All research ideas, modeling choices, theoretical claims, proof strategies, and experimental objectives were proposed by the authors, and generative AI was used only to help examine and realize these ideas provided by the authors. We did not use generative AI to independently originate the research questions or conclusions of this work. In addition, we used generative AI tools to help draft, revise, and polish parts of the manuscript. We reviewed all AI-assisted work, including mathematical arguments, simulation code, experimental outputs, and the manuscript. We take responsibility for the final content of this work, including text, claims, code, and other artifacts produced with the aid of generative AI.

REPRODUCIBILITY STATEMENT

The algorithm is specified in Section 3 and Algorithm 1, and the routing oracle and its implementation in Appendix B. All assumptions are stated in Sections 4.1 and 4.3, and complete proofs are given in Appendix C. Appendix D lists every setting of the experiments, including instance generation, hyperparameters, and baselines.

REFERENCES

Yasin Abbasi-Yadkori, Dávid Pál, and Csaba Szepesvári. Improved algorithms for linear stochastic bandits. *Advances in Neural Information Processing Systems*, 24, 2011.

Ravindra K. Ahuja, Thomas L. Magnanti, and James B. Orlin. *Network Flows: Theory, Algorithms, and Applications*. Prentice Hall, 1993.

Pranjal Awasthi, Kush Bhatia, Sreenivas Gollapudi, and Kostas Kollias. Congested bandits: Optimal routing via short-term resets. In *International Conference on Machine Learning*, pp. 1078–1100. PMLR, 2022.

Jørgen Bang-Jensen and Gregory Z. Gutin. *Digraphs: Theory, Algorithms and Applications*. Springer Monographs in Mathematics. Springer, 2 edition, 2009.

Hamsa Bastani, Mohsen Bayati, and Khashayar Khosravi. Mostly exploration-free algorithms for contextual bandits. *Management Science*, 67(3):1329–1349, 2021.

Avrim Blum, Eyal Even-Dar, and Katrina Ligett. Routing without regret: On convergence to Nash equilibria of regret-minimizing algorithms in routing games. In *Proceedings of the Twenty-Fifth Annual ACM Symposium on Principles of Distributed Computing*, pp. 45–52, 2006.

Stephen Boyd and Lieven Vandenberghe. *Convex Optimization*. Cambridge University Press, 2004.

Chao Chen, Karl Petty, Alexander Skabardonis, Pravin Varaiya, and Zhanfeng Jia. Freeway performance measurement system: Mining loop detector data. *Transportation Research Record*, 1748 (1):96–102, 2001.

Wei Chen, Yajun Wang, and Yang Yuan. Combinatorial multi-armed bandit: General framework and applications. In *International Conference on Machine Learning*, pp. 151–159. PMLR, 2013.

Qiwen Cui, Zhihan Xiong, Maryam Fazel, and Simon Du. Learning in congestion games with bandit feedback. *Advances in Neural Information Processing Systems*, 35:11009–11022, 2022.

Varsha Dani, Thomas P. Hayes, and Sham M. Kakade. Stochastic linear optimization under bandit feedback. In *Conference on Learning Theory*, pp. 355–366, 2008.

Yi Gai, Bhaskar Krishnamachari, and Rahul Jain. Combinatorial network optimization with unknown variables: Multi-armed bandits with linear rewards and individual observations. *IEEE/ACM Transactions on Networking*, 20(5):1466–1478, 2012.

Sreenivas Gollapudi, Kostas Kollias, Chinmay Maheshwari, and Manxi Wu. Online learning for traffic navigation in congested networks. In *International Conference on Algorithmic Learning Theory*, pp. 642–662. PMLR, 2023.

Geoffrey M. Guisewite and Panos M. Pardalos. Minimum concave-cost network flow problems: Applications, complexity, and algorithms. *Annals of Operations Research*, 25(1):75–99, 1990.

András György, Tamás Linder, Gábor Lugosi, and György Ottucsák. The on-line shortest path problem under partial monitoring. *Journal of Machine Learning Research*, 8(10), 2007.

Martin Jaggi. Revisiting Frank-Wolfe: Projection-free sparse convex optimization. In *International Conference on Machine Learning*, pp. 427–435. PMLR, 2013.

Sampath Kannan, Jamie H Morgenstern, Aaron Roth, Bo Waggoner, and Zhiwei Steven Wu. A smoothed analysis of the greedy algorithm for the linear contextual bandit problem. *Advances in Neural Information Processing Systems*, 31, 2018.

Johannes Kirschner and Andreas Krause. Information directed sampling and bandits with heteroscedastic noise. In *Conference on Learning Theory*, pp. 358–384. PMLR, 2018.

Johannes Kirschner and Andreas Krause. Stochastic bandits with context distributions. *Advances in Neural Information Processing Systems*, 32, 2019.

Branislav Kveton, Zheng Wen, Azin Ashkan, and Csaba Szepesvari. Tight regret bounds for stochastic combinatorial semi-bandits. In *Artificial Intelligence and Statistics*, pp. 535–543. PMLR, 2015.

Tor Lattimore, Koby Crammer, and Csaba Szepesvári. Optimal resource allocation with semi-bandit feedback. In *Proceedings of the 30th Conference on Uncertainty in Artificial Intelligence*, 2014.

Tor Lattimore, Koby Crammer, and Csaba Szepesvári. Linear multi-resource allocation with semi-bandit feedback. *Advances in Neural Information Processing Systems*, 28, 2015.

Larry J LeBlanc, Edward K Morlok, and William P Pierskalla. An efficient approach to solving the road network equilibrium traffic assignment problem. *Transportation Research*, 9(5):309–318, 1975.

Lihong Li, Wei Chu, John Langford, and Robert E Schapire. A contextual-bandit approach to personalized news article recommendation. In *Proceedings of the 19th International Conference on World Wide Web*, pp. 661–670, 2010.

Wenhao Li and Ningyuan Chen. Allocating divisible resources on arms with unknown and random rewards. In *Conference on Learning Theory*, pp. 2350–2351, 2023.

Hao Qin and Chicheng Zhang. Optimal and efficient contextual combinatorial semi-bandits with general function approximation. *arXiv preprint arXiv:2607.13686*, 2026.

Tim Roughgarden. Routing games. In Noam Nisan, Tim Roughgarden, Éva Tardos, and Vijay V. Vazirani (eds.), *Algorithmic Game Theory*, pp. 461–486. Cambridge University Press, 2007.

Kei Takemura, Shinji Ito, Daisuke Hatano, Hanna Sumita, Takuro Fukunaga, Naonori Kakimura, and Ken-ichi Kawarabayashi. Near-optimal regret bounds for contextual combinatorial semi-bandits with linear payoff functions. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, pp. 9791–9798, 2021.

Transportation Networks for Research Core Team. Transportation networks for research. `https://github.com/bstabler/TransportationNetworks`, 2016. Accessed September 1, 2026.

Joel A. Tropp. Freedman's inequality for matrix martingales. *Electronic Communications in Probability*, 16:262–270, 2011.

Chaoqi Wang, Ziyu Ye, Zhe Feng, Ashwinkumar Badanidiyuru Varadaraja, and Haifeng Xu. Follow-ups also matter: Improving contextual bandits via post-serving contexts. *Advances in Neural Information Processing Systems*, 36:12774–12796, 2023.

Zheng Wen, Branislav Kveton, and Azin Ashkan. Efficient learning in large-scale combinatorial semi-bandits. In *International Conference on Machine Learning*, pp. 1113–1122. PMLR, 2015.

A Detailed structure

A.1 Routing model in a congested network

Here we write down our setting in Section 2 in detail. It is similar to the non-atomic routing model of Roughgarden (2007). Throughout, $\mathbb{R}_+ = [0, \infty)$.

Let $G = (V, E)$ be a directed multigraph (Bang-Jensen & Gutin, 2009). We suppose that every edge $e \in E$ has a unique label with its tail and head well-defined by $\mathrm{tail}(e) \in V$ and $\mathrm{head}(e) \in V$, respectively. G may have parallel edges, but no loops.

A simple path of length $\ell \geq 1$ is a sequence of edges $p = (e_1, \ldots, e_\ell)$, whose associated vertex sequence $(v_0, \ldots, v_\ell)$ satisfies $\mathrm{tail}(e_j) = v_{j-1}$ and $\mathrm{head}(e_j) = v_j$, $j = 1, \ldots, \ell$, and $v_j \neq v_k$ for $0 \leq j < k \leq \ell$. The vertices v_0 and v_ℓ are the origin and destination of p, respectively. For an edge $e \in E$ and a path $p \in \mathcal{P}$, write $e \in p$ if e appears in the edge sequence defining p.

An origin–destination (OD) pair $(o, d) \in V \times V$ is called feasible if there exists at least one simple path from o to d. We define $\mathcal{I}$ as the set of all feasible OD pairs of G, and call its elements commodities. For each commodity $i = (o_i, d_i) \in \mathcal{I}$, let $\mathcal{P}_i$ be the set of all simple paths from o_i to d_i, and $\mathcal{P} = \bigcup_{i \in \mathcal{I}} \mathcal{P}_i$ be the set of all admissible paths in G.

A path flow of G is denoted as a vector $f = (f_p)_{p \in \mathcal{P}} \in \mathbb{R}_+^{\mathcal{P}}$, where $f_p \geq 0$ denotes the nonnegative amount of traffic assigned to the path p. Given a path flow f, its induced edge flow on $e \in E$ is denoted as $n_e(f) = \sum_{p: e \in p} f_p \geq 0$.

Denote $\mathcal{X}$ as a compact space of contexts. For each edge $e \in E$, let $c_e : \mathcal{X} \times \mathbb{R}_+ \to \mathbb{R}_+$ be its cost function, where $c_e(x, z)$ is the cost per unit flow in context x and edge flow z. For every fixed $x \in \mathcal{X}$, the function $z \mapsto c_e(x, z)$ is nonnegative and nondecreasing on $\mathbb{R}_+$. The affine specification in Section 2 is a special case.

A demand vector in G is denoted by $r = (r_i)_{i \in \mathcal{I}} \in \mathbb{R}_+^{\mathcal{I}}$ where $r_i \geq 0$ denotes the amount of demand of commodity i. Define the feasible flow set by

$$\mathcal{F}(r) = \left\{ f \in \mathbb{R}_+^{\mathcal{P}} : \sum_{p \in \mathcal{P}_i} f_p = r_i \quad \forall i \in \mathcal{I} \right\}.$$

Thus, the demand r_i of each commodity i must be exactly distributed among the paths in $\mathcal{P}_i$. In particular, every $f \in \mathcal{F}(r)$ satisfies $\sum_{p \in \mathcal{P}} f_p = \sum_{i \in \mathcal{I}} r_i$. We can see that $\mathcal{F}(r)$ is a nonempty compact convex polytope.

Remark 1 (Comparison to social objective) *The platform's criterion $C(f; x, u)$ counts only the cost of its controlled users, which matches a navigation platform that optimizes the travel times of its own users. The social cost $\sum_e (n_e(f) + u_e) c_e(x, n_e(f) + u_e)$ also counts the background users. Under affine costs, its conditional expectation equals $\sum_e (n_e + \mu_e)\{a_e + b_e(n_e + \mu_e)\}$ plus a term that does not depend on f whenever the conditional variance of background flow does not depend on the decision, so it is again a convex quadratic in the controlled loads. The optimistic reduction and the routing oracle therefore apply without change; we do not pursue the corresponding regret analysis here.*

A.2 Algorithm details

Updates. At $t = 1$, all sums in Eq. (4) and Eq. (5) are empty, which gives the initialization $V_1^u = \lambda_u I_{d_u}$, $V_{1,e}^c = \lambda_c I_{2d_c}$, and $h_{1,e}^u = h_{1,e}^c = 0$ in Algorithm 1. After round t, the statistics are updated incrementally:

$$\begin{aligned} V_{t+1}^u &= V_t^u + \phi_u(X_t)\phi_u(X_t)^\top, \quad & h_{t+1,e}^u &= h_{t,e}^u + \phi_u(X_t) U_{t,e} \quad (e \in E), \\ V_{t+1,e}^c &= V_{t,e}^c + \chi_{t,e}\chi_{t,e}^\top, & h_{t+1,e}^c &= h_{t,e}^c + \chi_{t,e} y_{t,e}^c \quad (e \in \mathcal{O}_t). \end{aligned} \tag{12}$$

The cost statistics of edges outside $\mathcal{O}_t$ stay unchanged. The weighted cost noise has scale one, so no further noise multiplier enters the cost radius $\alpha_{t,e}^c$.

Confidence sets and projection. The analysis uses the confidence sets

$$\mathcal{C}^{\gamma}_{t,e} = \{\gamma : \|\gamma - \widehat{\gamma}_{t,e}\|_{V^u_t} \le \alpha^u_t\}, \qquad \mathcal{C}^{\beta}_{t,e} = \{\beta : \|\beta - \widehat{\beta}_{t,e}\|_{V^c_{t,e}} \le \alpha^c_{t,e}\}.$$

Since ϕ_c contains a constant component, $\phi_c(X_t) \neq 0$, so A_t has full row rank and $G_{t,e}$ is positive definite.

The transformed 2-by-2 ellipse

$$\mathcal{C}^{\zeta}_{t,e} = \{\zeta : \|\zeta - \widehat{\zeta}_{t,e}\|_{G^{-1}_{t,e}} \le \alpha^c_{t,e}\}$$

contains $\zeta_{t,e}$ whenever $\beta_e \in \mathcal{C}^{\beta}_{t,e}$ (Lemma 11).

For comparison, the unconstrained ellipsoidal lower prediction is

$$\widehat{a}_{t,e} + z\widehat{b}_{t,e} - \alpha^c_{t,e}\sqrt{(G_{t,e})_{11} + 2z(G_{t,e})_{12} + z^2(G_{t,e})_{22}}.$$

It need not be affine, nonnegative, or nondecreasing in z. A common rectangle corner avoids flow-dependent reselection. This is computationally useful, but discards statistical dependence and can fail without compatibility.

B Detailed description of the routing oracle

This section describes the routing oracle Eq. (8), the subroutine that receives the surrogate model in round t and returns a controlled path flow. We show that it is a polynomial-size convex program that never enumerates paths; therefore, any convex programming method applies. Our implementation uses Frank–Wolfe. The main text and Theorem 1 assume an exact oracle. Appendix B.2 discusses computational error in practice and quantifies its effect on the regret.

B.1 Routing oracle as a polynomial-size convex program

In round t, the oracle receives the request vector R_t and, for every edge e, the optimistic coefficients $\underline{a}_{t,e}, \underline{b}_{t,e} \ge 0$ of Eq. (7) and the background-flow prediction $\widetilde{\mu}_{t,e} \ge 0$. The context X_t enters only through these inputs, which are computed before the oracle is called. Write

$$d_{t,e} = \underline{a}_{t,e} + \underline{b}_{t,e}\widetilde{\mu}_{t,e} \ge 0, \qquad F_t(n) = \sum_{e\in E}\big(\underline{b}_{t,e}n_e^2 + d_{t,e}n_e\big), \qquad n \in \mathbb{R}^E_+.$$

For every path flow f with edge loads $n(f) = (n_e(f))_{e\in E}$,

$$F_t(n(f)) = \sum_{e\in E} n_e(f)\,\widetilde{q}_{t,e}(n_e(f)) = M_t\widetilde{L}_t(f)$$

is the surrogate total cost of the controlled flow, where $\widetilde{q}_{t,e}(n) = \underline{a}_{t,e} + \underline{b}_{t,e}(n + \widetilde{\mu}_{t,e})$ is the surrogate unit cost and $\widetilde{L}_t(f)$ is the surrogate average cost.

The oracle returns

$$f_t \in \operatorname*{arg\,min}_{f\in\mathcal{F}(R_t)} F_t\big(n(f)\big), \tag{13}$$

which is exactly Eq. (8). Since M_t is fixed within a round, f_t also minimizes $\widetilde{L}_t$ over $\mathcal{F}(R_t)$.

The problem Eq. (13) has one variable per simple path, and $|\mathcal{P}_i|$ can be exponential in the size of the graph. For example, a chain of $k+1$ nodes with two parallel edges between consecutive nodes has $2k$ edges but 2^k simple paths between its endpoints. Enumerating and comparing paths is therefore not a viable implementation. The following lemma shows that this is unnecessary.

Lemma 2 *F_t is a separable convex quadratic, nondecreasing in each coordinate in $\mathbb{R}^E_+$, and $f \mapsto F_t(n(f))$ is convex in $\mathcal{F}(R_t)$.*

Proof Each summand $\underline{b}_{t,e}n_e^2 + d_{t,e}n_e$ has second derivative $2\underline{b}_{t,e} \ge 0$ and, for $n_e \ge 0$, first derivative $2\underline{b}_{t,e}n_e + d_{t,e} \ge 0$. The map $f \mapsto n(f)$ is linear, and a convex function composed with a linear map is convex. ∎

Lemma 2 rests on the nonnegativity of $\underline{a}_{t,e}$, $\underline{b}_{t,e}$, and $\widetilde{\mu}_{t,e}$, which the algorithm enforces by truncating at zero. They would not be guaranteed for a pointwise-optimistic cost over the projected ellipse, which need not be affine, nonnegative, or nondecreasing in the load (Appendix A.2).

Let $\mathcal{I}_t = \{i \in \mathcal{I} : R_{t,i} > 0\}$ be the active commodities in round t, and let $\mathsf{N} \in \{0, \pm 1\}^{V \times E}$ be the node–edge incidence matrix, with $\mathsf{N}_{v,e} = +1$ if v is the tail of e, $\mathsf{N}_{v,e} = -1$ if v is its head, and $\mathsf{N}_{v,e} = 0$ otherwise. This is well defined because the graph G has no loops, and the parallel edges are distinct columns. With a commodity-specific edge flow $h_i \in \mathbb{R}^E$ for each $i \in \mathcal{I}_t$, consider

$$\begin{aligned} \min_{h,\, n} \quad & F_t(n) \\ \text{subject to} \quad & \mathsf{N}h_i = R_{t,i}(\mathbf{e}_{o_i} - \mathbf{e}_{d_i}), && i \in \mathcal{I}_t, \\ & 0 \le h_{i,e} \le R_{t,i}, && i \in \mathcal{I}_t,\ e \in E, \\ & n_e = \textstyle\sum_{i \in \mathcal{I}_t} h_{i,e}, && e \in E. \end{aligned} \tag{14}$$

It has $(|\mathcal{I}_t| + 1)|E|$ variables, $|\mathcal{I}_t||V| + |E|$ sparse equality constraints, and $2|\mathcal{I}_t||E|$ bounds. The objective is a convex quadratic whose Hessian is $2\mathrm{diag}(\underline{b}_{t,e})$ in the n block and zero elsewhere. Hence Eq. (14) is a convex quadratic program whose size is polynomial in $|V|$, $|E|$, and $|\mathcal{I}_t| \le |V|(|V| - 1)$. The bounds $h_{i,e} \le R_{t,i}$ are not network capacities: every simple-path flow satisfies them, because a simple path uses each edge at most once. They only make the feasible set compact by excluding arbitrarily large circulations. Zero slopes and multiple minimizers are allowed.

Proposition 2 (Equivalence with the path program) *Fix a round t. The following hold.*

(i) Every $f \in \mathcal{F}(R_t)$ induces a feasible point of Eq. (14) with $n = n(f)$.

(ii) Every feasible point (h, n) of Eq. (14) can be converted, in time polynomial in $|V|$, $|E|$, and $|\mathcal{I}_t|$, into a simple-path flow $f \in \mathcal{F}(R_t)$ with $n(f) \le n$ componentwise, and hence $F_t(n(f)) \le F_t(n)$.

(iii) The optimal values of Eq. (14) and Eq. (13) coincide. If (h, n) is within $\epsilon \ge 0$ of the optimal value of Eq. (14), then the path flow from (ii) is within ϵ of the optimal value of Eq. (13).

Proof (i) Set $h_{i,e} = \sum_{p \in \mathcal{P}_i : e \in p} f_p$. Each simple o_i–d_i path contributes $+f_p$ to the net outflow at o_i, $-f_p$ at d_i, and zero at every other node, so $\mathsf{N}h_i = R_{t,i}(\mathbf{e}_{o_i} - \mathbf{e}_{d_i})$. Moreover $0 \le h_{i,e} \le \sum_{p \in \mathcal{P}_i} f_p = R_{t,i}$ and $\sum_i h_{i,e} = n_e(f)$.

(ii) Fix i and write $h = h_i$. While the support of h contains a directed cycle C, subtract $\min_{e \in C} h_e$ from every edge of C. This keeps $h \ge 0$ and $\mathsf{N}h$ unchanged, never increases a coordinate, and zeroes at least one edge. After at most $|E|$ cancellations it yields an acyclic flow $h' \le h$ with the same net outflows. Now, while the net outflow of h' at o_i is positive, start at o_i and follow edges with positive flow. At every node other than o_i and d_i inflow equals outflow, so the walk can continue until it reaches d_i, and acyclicity forbids revisiting a node. The walk is therefore a simple path $p \in \mathcal{P}_i$. Subtract $\min_{e \in p} h'_e$ along p and add this amount to f_p, which zeroes at least one edge and lowers the net outflow at o_i by the same amount. When the net outflow at o_i reaches zero, h' is an acyclic flow with zero net outflow at every node, so $h' = 0$. The amounts assigned to paths thus sum to $R_{t,i}$. This is the flow decomposition theorem (Ahuja et al., 1993), and each commodity needs at most $2|E|$ subtractions, each found in $O(|V| + |E|)$ time. Doing this for every commodity yields $f \in \mathcal{F}(R_t)$ with $n(f) = \sum_i h'_i \le \sum_i h_i = n$, and Lemma 2 gives $F_t(n(f)) \le F_t(n)$.

(iii) By (i), the optimal value of Eq. (14) is at most that of Eq. (13); by (ii), the reverse inequality holds. If $F_t(n)$ exceeds the common optimal value by at most ϵ, then so does $F_t(n(f)) \le F_t(n)$. ■

The decomposition in (ii) must be carried out before execution. An optimal edge solution may contain circulations on edges with $\underline{b}_{t,e} = d_{t,e} = 0$, which do not change the surrogate cost but would inflate the executed loads. Executing a simple-path flow guarantees $\sum_e n_{t,e} \le H_G M_t$ (Lemma 7), which the regret analysis uses.

Any convex quadratic programming solver can be applied to Eq. (14). Interior-point methods compute a feasible point within any prescribed objective error $\epsilon' > 0$ using a number of Newton steps that grows as $O(\sqrt{m} \log(m/\epsilon'))$, where $m = 2|\mathcal{I}_t||E|$ is the number of inequality constraints; each

step solves a sparse linear system of the size of Eq. (14) (Boyd & Vandenberghe, 2004, Chapter 11). Proposition 2(iii) then converts the solution into a simple-path flow with the same error guarantee. Because the error enters only logarithmically, high accuracy is inexpensive with this approach. Any other convex programming method can be used in the same way.

For our numerical experiments, we solve Eq. (13) by the Frank–Wolfe method, the classical choice for convex traffic-assignment programs (LeBlanc et al., 1975). We refer to Jaggi (2013) for the method and its guarantee. The standard duality-gap guarantee (Jaggi, 2013, Theorem 2) then bounds the number of iterations needed to reach accuracy ϵ_t in the average objective by $O(1 + \underline{b}_t^{\max} H_G M_t/\epsilon_t)$, where $\underline{b}_t^{\max} = \max_e \underline{b}_{t,e}$.

B.2 Computational error and the regret bound

In practice, implemented routing oracles stop at a positive tolerance. We now quantify the effect of this computational error on the regret.

Definition 1 (Accurate oracle) *For $\epsilon \geq 0$, a routing oracle is* ϵ-accurate at round t *if it returns $f_t \in \mathcal{F}(R_t)$ with $\widetilde{L}_t(f_t) \leq \min_{f \in \mathcal{F}(R_t)} \widetilde{L}_t(f) + \epsilon$.*

Remark 2 (Regret with an inexact oracle) *Suppose the conditions of Theorem 1 hold, except that at each round f_t is returned by an $\epsilon_t^{\mathrm{opt}}$-accurate oracle, with a deterministic tolerance schedule and a deterministic solver. Then, with probability at least $1 - \delta$,*

$$\mathit{Regret}_T \leq B_T^{\mathrm{hp}}(\kappa_*) + \sum_{t=1}^{T} \epsilon_t^{\mathrm{opt}},$$

where B_T^{hp} is the explicit bound of Proposition 3. Under the schedule $\epsilon_t^{\mathrm{opt}} = \epsilon_0 t^{-1/2}$ or $\epsilon_t^{\mathrm{opt}} = \epsilon_0 t^{-1}$, the added term is at most $2\epsilon_0\sqrt{T}$ or $\epsilon_0(1 + \log T)$, respectively, so the rate of Theorem 1 is unchanged.

A deterministic schedule and solver keep f_t $\mathcal{D}_t^0$-measurable, so the routing rule remains non-anticipating and Proposition 1 applies to the trajectory actually generated. The proof of Proposition 3 uses the optimality of f_t only in Lemma 14, through $\widetilde{L}_t(f_t) \leq \widetilde{L}_t(f_t^\star)$. Relaxing this to $\widetilde{L}_t(f_t) \leq \widetilde{L}_t(f_t^\star) + \epsilon_t^{\mathrm{opt}}$ adds $\epsilon_t^{\mathrm{opt}}$ to the one-step bound. Every other step uses only $f_t \in \mathcal{F}(R_t)$ and the measurability of f_t. The price of a decreasing tolerance is computational only.

C Proof of Theorem 1

Theorem 1 combines two results. The first is an explicit regret bound that holds on the event where $1 + \rho_{t,e}$ stays above a level κ (Proposition 3). The second is Proposition 1, which guarantees this event with $\kappa = \kappa_*$ under residual variation. Appendix C.1 states the explicit bound, Appendix C.2 collects standard results from the literature, Appendix C.3 proves the explicit bound, Appendix C.4 proves Proposition 1, and Appendix C.5 combines the two. Appendix C.6 shows that the compatibility event cannot be dropped.

Conventions. For positive definite A, write $\|x\|_A = \sqrt{x^\top A x}$. Throughout, $(\tau_0, \tau_1) \neq (0, 0)$, so that $\bar{v}(n) = \tau_0^2 + \tau_1^2/n \in (0, \infty)$ and $\omega(n) = 1/\bar{v}(n) = n/(\tau_0^2 n + \tau_1^2)$ are well defined for $n > 0$, and ω is nondecreasing on $(0, \infty)$. We write $\underline{\zeta}_{t,e} = (\underline{a}_{t,e}, \underline{b}_{t,e})^\top$, $\underline{c}_{t,e}(z) = \underline{a}_{t,e} + \underline{b}_{t,e} z$, $g_{t,e}(z) = c_e(X_t, z) - \underline{c}_{t,e}(z)$, $n_{t,e}^\star = n_e(f_t^\star)$, and $\Delta_t = L_t(f_t) - L_t(f_t^\star)$ for the one-round regret, so that $\mathrm{Regret}_T = \sum_{t=1}^T \Delta_t$. Since $M_t > 0$, every round routes positive flow on at least one edge.

C.1 An explicit regret bound on the compatibility event

Fix $\delta \in (0, 1)$ and write $\delta' = \delta/5$. Define the analysis budgets

$$\overline{Z}_T = \overline{m} + B_0 + \sigma\sqrt{2\log(|E|T/\delta')}, \qquad Q_T = B_0(1 + \overline{Z}_T), \qquad Q_{\max} = B_0(1 + \overline{m} + B_0),$$

$$\bar{\Gamma}_u(T) = d_u \log\left(1 + \frac{TB_\phi^2}{\lambda_u d_u}\right), \qquad \bar{\Gamma}_c(T) = 2d_c \log\left(1 + \frac{T\omega(\overline{m})B_\phi^2(1+\overline{Z}_T^2)}{2d_c\lambda_c}\right),$$

$c_u = 1 + B_\phi^2/\lambda_u$, and the deterministic radius budgets

$$\overline{\alpha}_u(T) = \sqrt{\lambda_u}B_\theta + \sigma\sqrt{\bar{\Gamma}_u(T) + 2\log(|E|/\delta')}, \qquad \overline{\alpha}_c(T) = \sqrt{\lambda_c}B_\theta + \sqrt{\bar{\Gamma}_c(T) + 2\log(|E|/\delta')}.$$

For $\kappa \in (0,1]$, the compatibility event is

$$\mathcal{E}_{\mathrm{sh}}(\kappa) = \big\{1 + \rho_{t,e} \geq \kappa \text{ for all } t \leq T,\ e \in E\big\}.$$

Proposition 3 (Explicit regret bound on the compatibility event) *Suppose Assumptions 1–3 hold with $(\tau_0, \tau_1) \neq (0,0)$. Let $\delta \in (0,1)$ and $\lambda_u, \lambda_c > 0$, and run Algorithm 1 with $\eta_{t,e} = \alpha^c_{t,e}(\delta')$. There is an event $\mathcal{E}_0$ with $\mathbb{P}(\mathcal{E}_0) \geq 1 - 4\delta'$ such that, for every $\kappa \in (0,1]$, on $\mathcal{E}_0 \cap \mathcal{E}_{\mathrm{sh}}(\kappa)$,*

$$\begin{aligned} \textit{Regret}_T \leq B_T^{\mathrm{hp}}(\kappa) := K(\kappa)\overline{\alpha}_c(T)\sqrt{2TH_G\bar{v}(\underline{m})|E|\bar{\Gamma}_c(T)} + \frac{|E|Q_T\bar{\Gamma}_c(T)}{\log 2} \\ + 2H_GB_0\overline{\alpha}_u(T)\sqrt{c_uT\bar{\Gamma}_u(T)} + H_GB_0\sigma\sqrt{2T\log(1/\delta')}. \end{aligned} \tag{15}$$

The four terms are the cost error in rounds of small leverage, the cost error in rounds of large leverage, the error of the background-flow prediction, and the martingale that transfers the cost gap from predicted to realized loads. The event $\mathcal{E}_0$ does not depend on κ.

Corollary 1 (Explicit logarithmic form) *Let $Z_0 = \overline{m} + B_0$ and define*

$$A_c = \frac{\omega(\overline{m})B_\phi^2}{2d_c\lambda_c}(1 + 2Z_0^2 + 4\sigma^2), \qquad A_u = \frac{B_\phi^2}{\lambda_u d_u}, \qquad \ell_T = \log\frac{|E|(T+1)(1+A_c)(1+A_u)}{\delta'}. \tag{16}$$

Then, for every $\kappa \in (0,1]$,

$$\begin{aligned} B_T^{\mathrm{hp}}(\kappa) \leq K(\kappa)\sqrt{\bar{v}(\underline{m})}\sqrt{8H_G|E|d_cT}\left\{\sqrt{\lambda_c}B_\theta\sqrt{\ell_T} + \sqrt{4d_c+2}\,\ell_T\right\} \\ + \frac{4|E|d_c}{\log 2}\left\{Q_{\max}\ell_T + \sqrt{2}B_0\sigma\ell_T^{3/2}\right\} \\ + 2H_GB_0\sqrt{c_ud_uT}\left\{\sqrt{\lambda_u}B_\theta\sqrt{\ell_T} + \sigma\sqrt{d_u+2}\,\ell_T\right\} + H_GB_0\sigma\sqrt{2T\ell_T}. \end{aligned} \tag{17}$$

Proof Since $|E|(T+1)/\delta' \geq 10$, $\ell_T > 1$. Moreover $\log(|E|T/\delta')$, $\log(|E|/\delta')$, and $\log(1/\delta')$ are all at most ℓ_T. Hence $\overline{Z}_T \leq Z_0 + \sigma\sqrt{2\ell_T}$, which gives $Q_T \leq Q_{\max} + \sqrt{2}B_0\sigma\sqrt{\ell_T}$ and, using $(x+y)^2 \leq 2x^2 + 2y^2$ and $\ell_T \geq 1$,

$$1 + \overline{Z}_T^2 \leq 1 + 2Z_0^2 + 4\sigma^2\ell_T \leq (1 + 2Z_0^2 + 4\sigma^2)\ell_T.$$

For $a \geq 0$ and $\ell \geq 1$, $1 + Ta\ell \leq (T+1)(1+a)\ell$. Therefore

$$\begin{aligned} &\bar{\Gamma}_c(T) \leq 2d_c\log(1 + TA_c\ell_T) \leq 2d_c\{\log(T+1) + \log(1+A_c) + \log\ell_T\} \leq 4d_c\ell_T, \\ &\bar{\Gamma}_u(T) \leq d_u\log\{(T+1)(1+A_u)\} \leq d_u\ell_T, \end{aligned}$$

where we used $\log(T+1) + \log(1+A_c) \leq \ell_T$ and $\log\ell_T \leq \ell_T$. Consequently $\overline{\alpha}_c(T) \leq \sqrt{\lambda_c}B_\theta + \sqrt{(4d_c+2)\ell_T}$ and $\overline{\alpha}_u(T) \leq \sqrt{\lambda_u}B_\theta + \sigma\sqrt{(d_u+2)\ell_T}$. Substituting these bounds into Eq. (15) term by term gives Eq. (17). ■

C.2 Standard tools

The following results are standard; we state them in the form used below.

Lemma 3 (Self-normalized ridge confidence; Abbasi-Yadkori et al., 2011, Theorems 1–2) *Let $\{\mathcal{A}_s\}_{s\geq 0}$ be a filtration. Let $v_s \in \mathbb{R}^d$ be $\mathcal{A}_{s-1}$-measurable and let $\nu_s \in \mathbb{R}$ be $\mathcal{A}_s$-measurable with $\mathbb{E}[\exp(x\nu_s) \mid \mathcal{A}_{s-1}] \leq \exp(x^2R^2/2)$ for all $x \in \mathbb{R}$ and some $R \geq 0$. Let $y_s = v_s^\top\theta + \nu_s$ with*

$\|\theta\|_2 \le S$. *For* $\lambda > 0$ *let* $V_t = \lambda I_d + \sum_{s<t} v_s v_s^\top$ *and* $\widehat{\theta}_t = V_t^{-1} \sum_{s<t} v_s y_s$. *Then for every* $\delta_1 \in (0,1)$, *with probability at least* $1-\delta_1$, *simultaneously for all* $t \ge 1$,

$$\|\widehat{\theta}_t - \theta\|_{V_t} \le R\sqrt{\log \frac{\det V_t}{\det(\lambda I_d)} + 2\log\frac{1}{\delta_1}} + \sqrt{\lambda} S.$$

If $R = 0$, then $\nu_s = 0$ almost surely, $V_t(\widehat{\theta}_t - \theta) = -\lambda\theta$, and the bound follows from $\|\lambda\theta\|_{V_t^{-1}} \le \sqrt{\lambda} S$.

Lemma 4 (Determinant identities; Abbasi-Yadkori et al., 2011, Lemmas 10–11) *Let* $V_1 = \lambda I_d$ *with* $\lambda > 0$ *and* $V_{s+1} = V_s + x_s x_s^\top$. *Then*

$$\log \frac{\det V_{s+1}}{\det V_s} = \log\bigl(1 + \|x_s\|^2_{V_s^{-1}}\bigr), \qquad \log \frac{\det V_t}{\det(\lambda I_d)} \le d \log\left(1 + \frac{\sum_{s<t}\|x_s\|_2^2}{d\lambda}\right).$$

The second bound follows from $\det V \le (\operatorname{tr} V/d)^d$.

Lemma 5 (Sub-Gaussian martingale concentration) *Let* $\{\mathcal{A}_t\}_{t\ge 0}$ *be a filtration and let* W_t *be* $\mathcal{A}_t$*-measurable with* $\mathbb{E}[\exp(xW_t) \mid \mathcal{A}_{t-1}] \le \exp(x^2 s^2/2)$ *for all* $x \in \mathbb{R}$ *and a deterministic* $s \ge 0$. *Then for every* $\delta_1 \in (0,1)$, $\mathbb{P}\bigl(\sum_{t=1}^T W_t \ge s\sqrt{2T\log(1/\delta_1)}\bigr) \le \delta_1$.

Proof By the tower property, $\mathbb{E}\exp(x\sum_{t\le T} W_t) \le \exp(Tx^2 s^2/2)$ for every $x > 0$. Markov's inequality with $x = \sqrt{2\log(1/\delta_1)/(Ts^2)}$ gives the claim. If $s = 0$, then $\sum_t W_t = 0$ almost surely. ∎

Lemma 6 (Time-uniform matrix boundary; Tropp, 2011) *Let* Υ_s *be random symmetric* $d \times d$ *matrices that are* $\mathcal{D}_{s+1}$*-measurable, and let* Θ_s *be symmetric and* $\mathcal{D}_s$*-measurable with* $\log \mathbb{E}[e^{\Upsilon_s} \mid \mathcal{D}_s] \preceq \Theta_s$, *where the conditional matrix moment generating functions are finite. Then for every* $\delta_1 \in (0,1)$,

$$\mathbb{P}\left\{\exists k \ge 1 :\ \lambda_{\max}\left(\sum_{s=1}^k (\Upsilon_s - \Theta_s)\right) \ge \log(d/\delta_1)\right\} \le \delta_1. \tag{18}$$

Proof Let $S_k = \sum_{s\le k}(\Upsilon_s - \Theta_s)$ and $M_k = \operatorname{tr}\exp(S_k)$, with $M_0 = d$. Since $S_{k-1} - \Theta_k$ is $\mathcal{D}_k$-measurable, Lieb's concavity theorem in the form $\mathbb{E}\operatorname{tr}\exp(H + \Upsilon) \le \operatorname{tr}\exp(H + \log \mathbb{E}e^{\Upsilon})$ (Tropp, 2011) and monotonicity of $A \mapsto \operatorname{tr}\exp(A)$ in the semidefinite order give $\mathbb{E}[M_k \mid \mathcal{D}_k] \le \operatorname{tr}\exp\{S_{k-1} - \Theta_k + \log \mathbb{E}[e^{\Upsilon_k} \mid \mathcal{D}_k]\} \le M_{k-1}$. By the tower property, (M_k) is a nonnegative supermartingale at the end-of-round times. Since $M_k \ge e^{\lambda_{\max}(S_k)}$, the maximal inequality $\mathbb{P}(\sup_k M_k \ge d/\delta_1) \le \delta_1$ proves Eq. (18). ∎

C.3 Proof of Proposition 3

Throughout this subsection we abbreviate

$$\alpha_t^u = \alpha_t^u(\delta'), \qquad \alpha_{t,e}^c = \alpha_{t,e}^c(\delta') = \eta_{t,e}, \qquad r_t^u = \alpha_t^u \,\|\phi_u(X_t)\|_{(V_t^u)^{-1}}.$$

C.3.1 Flow exposure and four events

Lemma 7 (Flow exposure) *For every* t *and* $f \in \mathcal{F}(R_t)$,

$$0 \le n_e(f) \le M_t, \qquad \sum_e n_e(f) \le H_G M_t, \qquad \sum_e n_e(f)^2 \le H_G M_t^2.$$

Moreover $0 \le \Delta_t \le H_G Q_{\max}$ *on every sample path.*

Proof A simple path contains each edge at most once, so $n_e(f) = \sum_{p \ni e} f_p \le \sum_p f_p = M_t$. Exchanging finite sums, $\sum_e n_e(f) = \sum_p |p| f_p \le H_G M_t$, and $\sum_e n_e(f)^2 \le M_t \sum_e n_e(f) \le H_G M_t^2$. For $0 \le n \le M_t$, Assumption 1 gives $0 \le q_{t,e}(n) = a_{t,e} + b_{t,e}(n + \mu_{t,e}) \le B_0(1 + \overline{m} + B_0) = Q_{\max}$. Hence $0 \le L_t(f) \le Q_{\max} \sum_e n_e(f)/M_t \le H_G Q_{\max}$, and $\Delta_t \ge 0$ because $f_t^\star$ minimizes L_t. ∎

We use the events

$$\mathcal{E}_u = \big\{\|\widehat{\gamma}_{t,e} - \gamma_e\|_{V_t^u} \le \alpha_t^u \text{ for all } t \ge 1,\ e \in E\big\},$$
$$\mathcal{E}_c = \big\{\beta_e \in \mathcal{C}_{t,e}^\beta \text{ for all } t \ge 1,\ e \in E\big\} = \big\{\|\widehat{\beta}_{t,e} - \beta_e\|_{V_{t,e}^c} \le \alpha_{t,e}^c \text{ for all } t \ge 1,\ e \in E\big\},$$
$$\mathcal{E}_z = \big\{Z_{t,e} \le \overline{Z}_T \text{ for all } t \le T,\ e \in E\big\},$$

and the transfer event $\mathcal{E}_{\mathrm{tr}}$ of Lemma 16. Set $\mathcal{E}_0 = \mathcal{E}_u \cap \mathcal{E}_c \cap \mathcal{E}_z \cap \mathcal{E}_{\mathrm{tr}}$.

Lemma 8 (Flow confidence) *$\mathbb{P}(\mathcal{E}_u) \ge 1 - \delta'$, and on $\mathcal{E}_u$, $|\widetilde{\mu}_{t,e} - \mu_{t,e}| \le r_t^u$ for all $t \ge 1$ and $e \in E$.*

Proof Fix e and apply Lemma 3 with $\mathcal{A}_s = \mathcal{D}_{s+1}^0$, $v_s = \phi_u(X_s)$, $\nu_s = \varepsilon_{s,e}$, $\theta = \gamma_e$, $S = B_\theta$, $R = \sigma$, $\lambda = \lambda_u$, and $\delta_1 = \delta'/|E|$. The covariate $\phi_u(X_s)$ is $\mathcal{D}_s^0$-measurable, and $\varepsilon_{s,e} = U_{s,e} - \mu_e(X_s)$ is revealed in round s, hence $\mathcal{D}_{s+1}^0$-measurable. By Assumption 2 and the tower property over $\mathcal{D}_s^0 \subseteq \mathcal{D}_s$, $\mathbb{E}[\exp(x\varepsilon_{s,e}) \mid \mathcal{D}_s^0] = \mathbb{E}[\mathbb{E}[\exp(x\varepsilon_{s,e}) \mid \mathcal{D}_s] \mid \mathcal{D}_s^0] \le \exp(x^2\sigma^2/2)$. By Eq. (1), $U_{s,e} = \phi_u(X_s)^\top \gamma_e + \varepsilon_{s,e}$, so Eq. (4) is the ridge estimator of the lemma, and its radius is $\alpha_t^u(\delta')$. A union bound over the $|E|$ edges gives $\mathbb{P}(\mathcal{E}_u) \ge 1 - \delta'$. On $\mathcal{E}_u$, Cauchy–Schwarz gives $|\widehat{\mu}_{t,e} - \mu_{t,e}| = |\phi_u(X_t)^\top(\widehat{\gamma}_{t,e} - \gamma_e)| \le r_t^u$. Since $\mu_{t,e} \ge 0$, truncation at zero cannot increase the error, so $|[\widehat{\mu}_{t,e}]_+ - \mu_{t,e}| \le |\widehat{\mu}_{t,e} - \mu_{t,e}|$. ∎

Lemma 9 (Cost confidence) $\mathbb{P}(\mathcal{E}_c) \ge 1 - \delta'$.

Proof Fix e and apply Lemma 3 with $\mathcal{A}_s = \mathcal{G}_{s+1}$ and

$$v_s = \chi_{s,e}\, \mathbb{I}\{e \in \mathcal{O}_s\}, \qquad \nu_s = \sqrt{\omega(n_{s,e})}\, \xi_{s,e}\, \mathbb{I}\{e \in \mathcal{O}_s\},$$

together with $\theta = \beta_e$, $S = B_\theta$, $R = 1$, $\lambda = \lambda_c$, and $\delta_1 = \delta'/|E|$. The covariate $\chi_{s,e} = \sqrt{\omega(n_{s,e})}\psi(X_s, Z_{s,e})$ and the indicator $\mathbb{I}\{e \in \mathcal{O}_s\}$ are functions of (X_s, f_s, U_s), hence $\mathcal{G}_s$-measurable. On a used edge, $\xi_{s,e} = Y_{s,e} - c_e(X_s, Z_{s,e})$ is revealed in round s, so ν_s is $\mathcal{G}_{s+1}$-measurable. By Assumption 3,

$$\mathbb{E}[\exp(x\nu_s) \mid \mathcal{G}_s] \le \exp\left\{\frac{x^2}{2}\,\omega(n_{s,e})\Big(\tau_0^2 + \frac{\tau_1^2}{n_{s,e}}\Big)\right\} = \exp(x^2/2)$$

on $\{e \in \mathcal{O}_s\}$, and trivially on its complement. Multiplying Eq. (2) by $\sqrt{\omega(n_{s,e})}$ gives $y_{s,e}^c = \chi_{s,e}^\top \beta_e + \nu_s$ on used edges, while unused rounds contribute nothing to Eq. (5). Hence Eq. (5) is the ridge estimator of the lemma, and its radius is $\alpha_{t,e}^c(\delta') = \eta_{t,e}$. A union bound over edges completes the proof. No independence between edges, or between ξ and ε, is used. ∎

Lemma 10 (Load and information budgets) *$\mathbb{P}(\mathcal{E}_z) \ge 1 - \delta'$. On $\mathcal{E}_z$, for every $e \in E$ and $t \le T + 1$, $\Gamma_{t,e}^c \le \bar{\Gamma}_c(T)$ and $\alpha_{t,e}^c \le \overline{\alpha}_c(T)$. On every sample path, $\Gamma_t^u \le \bar{\Gamma}_u(T)$ and $\alpha_t^u \le \overline{\alpha}_u(T)$ for $t \le T + 1$.*

Proof Let $x_T = \sigma\sqrt{2\log(|E|T/\delta')}$. If $\sigma > 0$, Assumption 2 and Chernoff's bound give $\mathbb{P}(\varepsilon_{t,e} > x_T \mid \mathcal{D}_t) \le \exp(-x_T^2/(2\sigma^2)) = \delta'/(|E|T)$. Taking expectations and a union bound over the $|E|T$ edge-rounds, $\varepsilon_{t,e} \le x_T$ for all $t \le T$ and e with probability at least $1 - \delta'$. If $\sigma = 0$, this holds almost surely. On this event, $Z_{t,e} = n_{t,e} + \mu_{t,e} + \varepsilon_{t,e} \le \overline{m} + B_0 + x_T = \overline{Z}_T$ by Lemma 7 and Assumption 1.

On $\mathcal{E}_z$, every used edge-round with $s \le T$ has $0 < n_{s,e} \le \overline{m}$ and $0 \le Z_{s,e} \le \overline{Z}_T$, where $Z_{s,e} \ge 0$ because $U_{s,e} \ge 0$. Monotonicity of ω and $\|\psi(x,z)\|_2^2 = (1+z^2)\|\phi_c(x)\|_2^2$ give $\|\chi_{s,e}\|_2^2 \le$

$\omega(\overline{m})B_\phi^2(1+\overline{Z}_T^2)$. Lemma 4 in dimension $2d_c$, with at most T updates before round $T+1$, gives $\Gamma^c_{T+1,e} \le \bar{\Gamma}_c(T)$. Log-determinants are nondecreasing in t, so the same bound holds for all $t \le T+1$, and Eq. (9) gives $\alpha^c_{t,e}(\delta') \le \overline{\alpha}_c(T)$. The flow statements follow in the same way from $\|\phi_u(X_s)\|_2 \le B_\phi$, without any event. ■

C.3.2 Geometry of the optimistic reduction

Lemma 11 (Exact image) *Let $V \succ 0$, let $A \in \mathbb{R}^{2\times 2d_c}$ have full row rank, and let $G = AV^{-1}A^\top$. For every $r \ge 0$ and $\widehat{\beta}$, $\{A\beta : \|\beta - \widehat{\beta}\|_V \le r\} = \{\zeta : \|\zeta - A\widehat{\beta}\|_{G^{-1}} \le r\}$. In particular, $\|Aw\|_{G^{-1}} \le \|w\|_V$ for every w.*

Proof For $y \in \mathbb{R}^2$, consider $\min\{w^\top V w : Aw = y\}$. The vector $w_0 = V^{-1}A^\top G^{-1}y$ is feasible. Every feasible w is $w_0 + w'$ with $Aw' = 0$, and $w_0^\top V w' = y^\top G^{-1}Aw' = 0$. Hence $w^\top V w = w_0^\top V w_0 + w'^\top V w' \ge y^\top G^{-1} y$, with equality at w_0. Taking $y = \zeta - A\widehat{\beta}$ proves both inclusions. ■

Since ϕ_c contains a constant component, $\phi_c(X_t) \ne 0$, so A_t has full row rank and $G_{t,e} \succ 0$. Lemma 11 therefore proves Eq. (6).

Lemma 12 (Directional rectangle comparison) *Let $G \in \mathbb{R}^{2\times 2}$ be positive definite with $\rho = G_{12}/\sqrt{G_{11}G_{22}}$. For every $z \ge 0$, with $h_z = (1,z)^\top$,*

$$\sqrt{G_{11}} + z\sqrt{G_{22}} \le \sqrt{\frac{2}{1+\rho}}\,\|h_z\|_G,$$

with equality at $z = \sqrt{G_{11}/G_{22}}$. This proves Eq. (10).

Proof Positive definiteness gives $|\rho| < 1$. Put $u = \sqrt{G_{11}} \ge 0$ and $v = z\sqrt{G_{22}} \ge 0$, so that $\|h_z\|_G^2 = u^2 + 2\rho uv + v^2$. Expanding,

$$u^2 + 2\rho uv + v^2 - \frac{1+\rho}{2}(u+v)^2 = \frac{1-\rho}{2}(u-v)^2 \ge 0.$$

Dividing by $(1+\rho)/2 > 0$ and taking square roots gives the inequality, with equality if and only if $u = v$. ■

Lemma 13 (Greatest convex minorant) *Let $G \succ 0$, $\eta \ge 0$, $h_z = (1,z)^\top$, and $\ell(z) = \widehat{a} + \widehat{b}z - \eta\|h_z\|_G$. Let $\ell^\star(z) = (\widehat{a} - \eta\sqrt{G_{11}}) + (\widehat{b} - \eta\sqrt{G_{22}})z$. Then $\ell^\star \le \ell$ on $[0,\infty)$, and every convex $\widetilde{c} : [0,\infty) \to \mathbb{R}$ with $\widetilde{c} \le \ell$ satisfies $\widetilde{c} \le \ell^\star$.*

Proof The triangle inequality gives $\|h_z\|_G \le \|(1,0)^\top\|_G + z\|(0,1)^\top\|_G = \sqrt{G_{11}} + z\sqrt{G_{22}}$, so $\ell^\star \le \ell$. Let $\widetilde{c}$ be convex with $\widetilde{c} \le \ell$, and fix $z \ge 0$. For every $w > z$, convexity gives $\widetilde{c}(z) \le (1 - z/w)\widetilde{c}(0) + (z/w)\widetilde{c}(w) \le (1 - z/w)\ell(0) + (z/w)\ell(w)$. Since $\|h_w\|_G/w \to \sqrt{G_{22}}$, we have $\ell(w)/w \to \widehat{b} - \eta\sqrt{G_{22}}$ as $w \to \infty$, and hence $\widetilde{c}(z) \le \ell(0) + z(\widehat{b} - \eta\sqrt{G_{22}}) = \ell^\star(z)$. ■

Proof of Lemma 1 Write $\zeta = \zeta_{t,e}$, $\widehat{\zeta} = \widehat{\zeta}_{t,e}$, $\underline{\zeta} = \underline{\zeta}_{t,e}$, $G = G_{t,e}$, $\eta = \eta_{t,e}$, and $h_z = (1,z)^\top$. If $\beta_e \in \mathcal{C}^\beta_{t,e}$, Lemma 11 gives $\|\widehat{\zeta} - \zeta\|_{G^{-1}} \le \eta$, and hence $|\widehat{\zeta}_j - \zeta_j| \le \eta\sqrt{G_{jj}}$ for $j = 1,2$.

Optimism. By Eq. (7), $\underline{\zeta}_j = [\widehat{\zeta}_j - \eta\sqrt{G_{jj}}]_+$. Since $\widehat{\zeta}_j - \eta\sqrt{G_{jj}} \le \zeta_j$ and $\zeta_j \ge 0$, we get $0 \le \underline{\zeta}_j \le \zeta_j$. Then $g_{t,e}(z) = h_z^\top(\zeta - \underline{\zeta}) \ge 0$ for $z \ge 0$.

Gap. Whether or not the truncation is active, $\underline{\zeta}_j \ge \widehat{\zeta}_j - \eta\sqrt{G_{jj}}$, so $\zeta_j - \underline{\zeta}_j \le (\zeta_j - \widehat{\zeta}_j) + \eta\sqrt{G_{jj}}$ for $j = 1,2$. Multiplying by the nonnegative entries of h_z and summing,

$$g_{t,e}(z) \le h_z^\top(\zeta - \widehat{\zeta}) + \eta\big(\sqrt{G_{11}} + z\sqrt{G_{22}}\big) \le \eta\|h_z\|_G + \eta\sqrt{\tfrac{2}{\kappa}}\,\|h_z\|_G = K(\kappa)\,\eta\,\|h_z\|_G.$$

The first term uses Cauchy–Schwarz, $|h_z^\top(\zeta - \widehat{\zeta})| \le \|h_z\|_G\|\zeta - \widehat{\zeta}\|_{G^{-1}}$, and the second uses Lemma 12 with $1 + \rho_{t,e} \ge \kappa$. Finally $A_t^\top h_z = \psi(X_t, z)$ gives $\|h_z\|_G = \|\psi(X_t,z)\|_{(V^c_{t,e})^{-1}}$. The bound $g_{t,e}(z) \le B_0(1+z)$ follows from $\underline{c}_{t,e}(z) \ge 0$, which holds on every sample path, and $c_e(X_t,z) \le B_0(1+z)$. ■

C.3.3 ONE-STEP DECOMPOSITION AND THE FLOW TERM

Recall the surrogate average cost $\widetilde{L}_t(f) = M_t^{-1} \sum_e n_e(f) \widetilde{q}_{t,e}(n_e(f))$ from Appendix B.1, which f_t minimizes over $\mathcal{F}(R_t)$, and let

$$D_t^c = \frac{1}{M_t} \sum_e n_{t,e}\, g_{t,e}(n_{t,e} + \mu_{t,e}).$$

Lemma 14 (One-step regret) *On $\mathcal{E}_u \cap \mathcal{E}_c$, for every t, $\Delta_t \le D_t^c + 2H_G B_0 r_t^u$.*

Proof For every $n \ge 0$,

$$\begin{aligned} q_{t,e}(n) - \widetilde{q}_{t,e}(n) &= (a_{t,e} - \underline{a}_{t,e}) + (b_{t,e} - \underline{b}_{t,e})(n + \mu_{t,e}) + \underline{b}_{t,e}(\mu_{t,e} - \widetilde{\mu}_{t,e}) \\ &= g_{t,e}(n + \mu_{t,e}) + \underline{b}_{t,e}(\mu_{t,e} - \widetilde{\mu}_{t,e}). \end{aligned}$$

Since f_t minimizes $\widetilde{L}_t$ over $\mathcal{F}(R_t)$, $\widetilde{L}_t(f_t) \le \widetilde{L}_t(f_t^\star)$. Therefore

$$\begin{aligned} \Delta_t &\le \big[L_t(f_t) - \widetilde{L}_t(f_t)\big] + \big[\widetilde{L}_t(f_t^\star) - L_t(f_t^\star)\big] \\ &= D_t^c + \frac{1}{M_t} \sum_e n_{t,e} \underline{b}_{t,e}(\mu_{t,e} - \widetilde{\mu}_{t,e}) - \frac{1}{M_t} \sum_e n_{t,e}^\star \Big\{ g_{t,e}(n_{t,e}^\star + \mu_{t,e}) + \underline{b}_{t,e}(\mu_{t,e} - \widetilde{\mu}_{t,e}) \Big\}. \end{aligned}$$

On $\mathcal{E}_c$, Lemma 1 gives $g_{t,e} \ge 0$ and $0 \le \underline{b}_{t,e} \le b_{t,e} \le B_0$. Dropping the nonpositive comparator gap and using Lemma 8 on $\mathcal{E}_u$,

$$\Delta_t \le D_t^c + \frac{B_0 r_t^u}{M_t} \sum_e \big(n_{t,e} + n_{t,e}^\star\big) \le D_t^c + 2H_G B_0 r_t^u,$$

where the last step applies Lemma 7 to both f_t and $f_t^\star$. ∎

Lemma 15 (Cumulative flow error) *On every sample path, $\sum_{t=1}^T r_t^u \le \overline{\alpha}_u(T) \sqrt{c_u T \bar{\Gamma}_u(T)}$.*

Proof Let $s_t = \|\phi_u(X_t)\|^2_{(V_t^u)^{-1}} \in [0, B_\phi^2/\lambda_u]$. Since $\log(1+s) \ge s/(1+s)$ for $s \ge 0$, we have $s_t \le (1+s_t)\log(1+s_t) \le c_u \log(1+s_t)$. Lemma 4 gives $\sum_{t \le T} \log(1+s_t) = \Gamma_{T+1}^u \le \bar{\Gamma}_u(T)$. By Lemma 10, $\alpha_t^u \le \overline{\alpha}_u(T)$, and Cauchy–Schwarz gives $\sum_t r_t^u \le \overline{\alpha}_u(T) \sqrt{T \sum_t s_t}$. ∎

C.3.4 TRANSFER TO REALIZED LOADS

Define

$$\widehat{D}_t^c = \frac{1}{M_t} \sum_e n_{t,e}\, g_{t,e}(Z_{t,e}), \qquad J_t = \sum_e \frac{n_{t,e}}{M_t} (b_{t,e} - \underline{b}_{t,e}) \varepsilon_{t,e}.$$

Because the algorithm does not bound $\underline{b}_{t,e}$ from above, we also use the analysis-only truncation

$$\underline{b}^\circ_{t,e} = \min\{\underline{b}_{t,e}, B_0\}, \qquad J_t^\circ = \sum_e \frac{n_{t,e}}{M_t} (b_{t,e} - \underline{b}^\circ_{t,e}) \varepsilon_{t,e}.$$

The learner never computes $\underline{b}^\circ_{t,e}$.

Lemma 16 (Affine transfer) *On every sample path, $D_t^c = \widehat{D}_t^c - J_t$, which is the affine transfer in the proof sketch of Section 4.4. There is an event $\mathcal{E}_{\mathrm{tr}}$ with $\mathbb{P}(\mathcal{E}_{\mathrm{tr}}) \ge 1 - \delta'$ on which $-\sum_{t=1}^T J_t^\circ \le H_G B_0 \sigma \sqrt{2T \log(1/\delta')}$. On $\mathcal{E}_c$, $J_t^\circ = J_t$ for all t. Consequently, on $\mathcal{E}_c \cap \mathcal{E}_{\mathrm{tr}}$,*

$$\sum_{t=1}^T D_t^c \le \sum_{t=1}^T \widehat{D}_t^c + H_G B_0 \sigma \sqrt{2T \log(1/\delta')}.$$

Proof Each $g_{t,e}$ is affine with slope $b_{t,e} - \underline{b}_{t,e}$, and $Z_{t,e} = (n_{t,e} + \mu_{t,e}) + \varepsilon_{t,e}$. Hence $g_{t,e}(Z_{t,e}) = g_{t,e}(n_{t,e} + \mu_{t,e}) + (b_{t,e} - \underline{b}_{t,e})\varepsilon_{t,e}$, and multiplying by $n_{t,e}/M_t$ and summing gives the identity.

Let $w_{t,e} = n_{t,e}(b_{t,e} - \underline{b}^{\circ}_{t,e})/M_t$. These weights are $\mathcal{D}_t$-measurable, and since $b_{t,e}, \underline{b}^{\circ}_{t,e} \in [0, B_0]$ on every sample path, Lemma 7 gives $w_t := \sum_e |w_{t,e}| \le H_G B_0$. In particular $\mathbb{E}[J^{\circ}_t \mid \mathcal{D}_t] = \mathbb{E}[J_t \mid \mathcal{D}_t] = 0$ by Assumption 2. If $w_t > 0$, put $p_e = |w_{t,e}|/w_t$. The conditional Hölder inequality and Assumption 2 give, for every $x \in \mathbb{R}$,

$$\mathbb{E}\Big[e^{-x\sum_e w_{t,e}\varepsilon_{t,e}} \,\Big|\, \mathcal{D}_t\Big] \le \prod_{e:p_e>0} \Big(\mathbb{E}\big[e^{-xw_t \operatorname{sgn}(w_{t,e})\varepsilon_{t,e}} \mid \mathcal{D}_t\big]\Big)^{p_e} \le \exp\big(\tfrac{1}{2}x^2 w_t^2 \sigma^2\big) \le \exp\big(\tfrac{1}{2}x^2 H_G^2 B_0^2 \sigma^2\big),$$

and the case $w_t = 0$ is trivial. Thus $-J^{\circ}_t$ is conditionally $H_G B_0 \sigma$-sub-Gaussian given $\mathcal{D}_t$. It is $\mathcal{D}_{t+1}$-measurable, because $\varepsilon_{t,e}$ is revealed in round t. Lemma 5 with $\mathcal{A}_t = \mathcal{D}_{t+1}$ and $\delta_1 = \delta'$ defines $\mathcal{E}_{\mathrm{tr}}$. No independence across edges is used, and no confidence event is conditioned on.

On $\mathcal{E}_c$, Lemma 1 gives $\underline{b}_{t,e} \le b_{t,e} \le B_0$, so $\underline{b}^{\circ}_{t,e} = \underline{b}_{t,e}$ and $J^{\circ}_t = J_t$. Combining with the identity proves the last display. ∎

C.3.5 Realized-design potentials

For $e \in \mathcal{O}_t$ define

$$u_{t,e} = \|\chi_{t,e}\|_{(V^c_{t,e})^{-1}}, \qquad v_{t,e} = \frac{n_{t,e}}{M_t\sqrt{\omega(n_{t,e})}},$$

and set $u_{t,e} = v_{t,e} = 0$ for $e \notin \mathcal{O}_t$.

Lemma 17 (Potentials) *On every sample path, for every t and e,*

$$\sum_e v^2_{t,e} \le H_G \bar{v}(\underline{m}), \qquad \sum_{t\le T:\, u_{t,e}\le 1} u^2_{t,e} \le 2\Gamma^c_{T+1,e}, \qquad \big|\{t \le T : u_{t,e} > 1\}\big| \le \frac{\Gamma^c_{T+1,e}}{\log 2}.$$

Proof Since $1/\omega(n) = \tau_0^2 + \tau_1^2/n$, Lemma 7 gives $\sum_e v^2_{t,e} = M_t^{-2}\sum_e(\tau_0^2 n^2_{t,e} + \tau_1^2 n_{t,e}) \le H_G(\tau_0^2 + \tau_1^2/M_t) \le H_G\bar{v}(\underline{m})$. By Lemma 4, $\Gamma^c_{t+1,e} - \Gamma^c_{t,e} = \log(1 + u^2_{t,e})$ for $e \in \mathcal{O}_t$, and $\Gamma^c_{t+1,e} = \Gamma^c_{t,e}$ otherwise. For $x \in [0,1]$, $x \le 2\log(1+x)$, which gives the second bound. Each round with $u_{t,e} > 1$ increases $\Gamma^c_{\cdot,e}$ by more than $\log 2$, which gives the third. ∎

Lemma 18 (Cumulative realized cost gap) *For every $\kappa \in (0,1]$, on $\mathcal{E}_c \cap \mathcal{E}_z \cap \mathcal{E}_{\mathrm{sh}}(\kappa)$,*

$$\sum_{t=1}^{T} \widehat{D}^c_t \le K(\kappa)\overline{\alpha}_c(T)\sqrt{2TH_G\bar{v}(\underline{m})|E|\bar{\Gamma}_c(T)} + \frac{|E|Q_T\bar{\Gamma}_c(T)}{\log 2}.$$

Proof Only used edges contribute to $\widehat{D}^c_t$. For $e \in \mathcal{O}_t$ and $t \le T$, $\|\psi(X_t, Z_{t,e})\|_{(V^c_{t,e})^{-1}} = u_{t,e}/\sqrt{\omega(n_{t,e})}$. Lemma 1 with $\eta_{t,e} = \alpha^c_{t,e}$ and Lemma 10 therefore give

$$\frac{n_{t,e}}{M_t} g_{t,e}(Z_{t,e}) \le K(\kappa)\alpha^c_{t,e} v_{t,e}u_{t,e} \le K(\kappa)\overline{\alpha}_c(T) v_{t,e}u_{t,e}.$$

Small leverage. By Cauchy–Schwarz and Lemma 17,

$$\sum_{t\le T}\sum_{e:u_{t,e}\le 1} v_{t,e}u_{t,e} \le \sqrt{\sum_{t\le T}\sum_e v^2_{t,e}}\sqrt{\sum_e \sum_{t\le T:u_{t,e}\le 1} u^2_{t,e}} \le \sqrt{TH_G\bar{v}(\underline{m})\cdot 2\textstyle\sum_e \Gamma^c_{T+1,e}},$$

and $\sum_e \Gamma^c_{T+1,e} \le |E|\bar{\Gamma}_c(T)$ on $\mathcal{E}_z$.

Large leverage. If $u_{t,e} > 1$, use instead $n_{t,e}/M_t \le 1$ and $g_{t,e}(Z_{t,e}) \le B_0(1 + Z_{t,e}) \le Q_T$ on $\mathcal{E}_z$. By Lemma 17, there are at most $\sum_e \Gamma^c_{T+1,e}/\log 2 \le |E|\bar{\Gamma}_c(T)/\log 2$ such edge-rounds. Adding the two parts proves the claim. ∎

C.3.6 COMPLETION OF THE PROOF

Proof of Proposition 3 By Lemmas 8, 9, 10, and 16, each of $\mathcal{E}_u, \mathcal{E}_c, \mathcal{E}_z, \mathcal{E}_{\mathrm{tr}}$ fails with probability at most δ', so a union bound gives $\mathbb{P}(\mathcal{E}_0) \geq 1 - 4\delta'$ without any independence assumption. All four events are constructed under the original probability law and do not involve κ. Fix $\kappa \in (0, 1]$ and work on $\mathcal{E}_0 \cap \mathcal{E}_{\mathrm{sh}}(\kappa)$. Lemmas 14 and 15 give

$$\mathrm{Regret}_T = \sum_{t=1}^{T} \Delta_t \leq \sum_{t=1}^{T} D_t^c + 2H_G B_0 \overline{\alpha}_u(T) \sqrt{c_u T \bar{\Gamma}_u(T)}.$$

Lemma 16 bounds the first sum by $\sum_t \widehat{D}_t^c + H_G B_0 \sigma \sqrt{2T \log(1/\delta')}$, and Lemma 18 bounds $\sum_t \widehat{D}_t^c$. Combining these gives Eq. (15). ∎

C.4 PROOF OF PROPOSITION 1

We prove a version for a general $\lambda_c > 0$ and then specialize to Eq. (11). Here and in Proposition 1, a non-anticipating routing rule chooses $f_t \in \mathcal{F}(R_t)$ from the history, (X_t, R_t), and possibly independent internal randomness, which is included in $\mathcal{D}_t$; Assumptions 2 and 4 are required for the process that the rule generates.

C.4.1 CONSTANTS

For $\delta \in (0, 1)$ and $\lambda_c > 0$, define

$$Z_0 = \overline{m} + B_0, \qquad h = \omega(\overline{m}) B_\phi^2, \qquad \ell_{\mathrm{sh}} = \log \frac{4d_c|E|}{\delta}, \qquad x_\delta = \frac{h\ell_{\mathrm{sh}}}{\lambda_c}.$$

Every round uses at least one edge, and a σ-sub-Gaussian variable has conditional variance at most σ^2. Hence Assumption 4 implies $0 < v_* \leq \sigma^2$, and in particular $\sigma > 0$. Set

$$c_0 = \frac{v_*}{1 + Z_0^2 + \sigma^2},$$

and let

$$r = 2\sigma \sqrt{\log \frac{4(1 + 2Z_0^2 + 7\sigma^2)}{c_0}}, \qquad K_* = 1 + (Z_0 + r)^2, \qquad a_* = (1 - e^{-1})c_0/2.$$

With $y = r^2/(2\sigma^2)$ and $ye^{-y/2} \leq 2/e$, this choice gives $2(1 + 2Z_0^2 + 2r^2 + 4\sigma^2)e^{-r^2/(2\sigma^2)} \leq 2(1 + 2Z_0^2 + 7\sigma^2)e^{-y/2} = c_0/2$. Note that $0 < a_* < 1$, since $c_0 < 1$. Finally let

$$c_- = \frac{a_*}{1 + K_* x_\delta}, \qquad C_+ = 2(1 + Z_0^2) + 4\sigma^2 + 8\sigma^2 x_\delta, \qquad \kappa_\delta = \frac{2c_-}{C_+ + c_-} \in (0, 1).$$

Under Eq. (11), $x_\delta = 1/\lambda_0$, and κ_δ equals the constant

$$\kappa_* = \frac{2c_*^-}{C_*^+ + c_*^-}, \qquad c_*^- = \frac{a_*}{1 + K_*/\lambda_0}, \qquad C_*^+ = 2(1 + Z_0^2) + 4\sigma^2 + 8\sigma^2/\lambda_0, \tag{19}$$

which depends only on Z_0, σ, v_*, and λ_0, and increases with v_* because c_0 increases and r decreases with v_*. These constants are used only in the analysis and are deliberately conservative.

C.4.2 CONDITIONAL LOAD GEOMETRY

Fix an edge e and suppress its index where no confusion arises. Write

$$\varphi_s = \phi_c(X_s), \qquad W_s = \omega(n_{s,e}) \varphi_s \varphi_s^\top \mathbb{I}\{e \in \mathcal{O}_s\},$$
$$S_t = \sum_{s<t} W_s, \qquad \mathsf{D}_t = \mathrm{diag}(\lambda_c I + S_t,\ \lambda_c I + S_t),$$

so that W_s is $\mathcal{D}_s$-measurable and $0 \preceq W_s \preceq hI$. Let $m_s = n_{s,e} + \mu_e(X_s) \in [0, Z_0]$, so that $Z_{s,e} = m_s + \varepsilon_{s,e}$.

Lemma 19 (Conditional load geometry and truncation) *Under Assumptions 1, 2, and 4, on every used edge-round,*

$$\mathbb{E}\left[\begin{pmatrix} 1 & Z_{s,e} \\ Z_{s,e} & Z_{s,e}^2 \end{pmatrix} \mathbb{I}\{|\varepsilon_{s,e}| \le r\} \,\middle|\, \mathcal{D}_s\right] \succeq \frac{c_0}{2} I_2.$$

Proof Let $v_s = \mathbb{E}[\varepsilon_{s,e}^2 \mid \mathcal{D}_s] \in [v_*, \sigma^2]$. Without truncation, the conditional expectation equals $\left(\begin{smallmatrix} 1 & m_s \\ m_s & m_s^2+v_s \end{smallmatrix}\right)$, whose determinant is $v_s \ge v_*$ and whose trace is at most $1 + Z_0^2 + \sigma^2$. Its smaller eigenvalue is at least the determinant divided by the trace, hence at least c_0. The removed part $\mathbb{E}[\left(\begin{smallmatrix} 1 & Z \\ Z & Z^2 \end{smallmatrix}\right)\mathbb{I}\{|\varepsilon| > r\} \mid \mathcal{D}_s]$ is positive semidefinite, so its operator norm is at most its trace $\mathbb{E}[(1 + Z^2)\mathbb{I}\{|\varepsilon| > r\} \mid \mathcal{D}_s]$. Using $Z^2 \le 2Z_0^2 + 2\varepsilon^2$ and the sub-Gaussian tails

$$\begin{aligned} \mathbb{P}(|\varepsilon_{s,e}| > x \mid \mathcal{D}_s) &\le 2e^{-x^2/(2\sigma^2)} \qquad (x \ge 0), \\ \mathbb{E}[\varepsilon_{s,e}^2 \mathbb{I}\{|\varepsilon_{s,e}| > r\} \mid \mathcal{D}_s] &= r^2 \mathbb{P}(|\varepsilon_{s,e}| > r \mid \mathcal{D}_s) + \int_r^\infty 2x\, \mathbb{P}(|\varepsilon_{s,e}| > x \mid \mathcal{D}_s)\, dx \\ &\le 2(r^2 + 2\sigma^2) e^{-r^2/(2\sigma^2)}, \end{aligned}$$

the trace is at most $2(1 + 2Z_0^2 + 2r^2 + 4\sigma^2)e^{-r^2/(2\sigma^2)} \le c_0/2$. Subtracting proves the claim. No symmetry of the residual is needed. ∎

C.4.3 Time-uniform comparison of the realized design

Lemma 20 (Design sandwich) *Under Assumptions 1, 2, and 4, for any non-anticipating routing rule using the weighted cost design, with probability at least $1 - \delta$, simultaneously for every $e \in E$ and $t \ge 1$,*

$$c_- \mathsf{D}_{t,e} \preceq V_{t,e}^c \preceq C_+ \mathsf{D}_{t,e}. \tag{20}$$

Proof Fix e. Since $\psi(X_s, z) = (1, z)^\top \otimes \varphi_s$, the cost design is $V_{t,e}^c = \lambda_c I + \sum_{s<t} \omega(n_{s,e}) \mathbb{I}\{e \in \mathcal{O}_s\} \psi\psi^\top$ with $\omega(n_{s,e})\mathbb{I}\{e \in \mathcal{O}_s\}\psi(X_s, Z_{s,e})\psi(X_s, Z_{s,e})^\top = \left(\begin{smallmatrix} 1 & Z_{s,e} \\ Z_{s,e} & Z_{s,e}^2 \end{smallmatrix}\right) \otimes W_s$.

Lower bound. Let $\Omega_s = \left(\begin{smallmatrix} 1 & Z_{s,e} \\ Z_{s,e} & Z_{s,e}^2 \end{smallmatrix}\right)\mathbb{I}\{|\varepsilon_{s,e}| \le r\} \otimes W_s$ and $\bar{h} = hK_*$. Since $|Z_{s,e}| \le Z_0 + r$ on the truncation event, $0 \preceq \Omega_s \preceq \bar{h} I_{2d_c}$. Lemma 19 and monotonicity of the Kronecker product with a positive semidefinite factor give $\mathbb{E}[\Omega_s \mid \mathcal{D}_s] \succeq (c_0/2)\mathrm{diag}(W_s, W_s)$; on unused rounds both sides vanish. For $x \in [0, 1]$, $e^{-x} \le 1 - (1 - e^{-1})x$ by convexity. Applying this to the eigenvalues of $\Omega_s/\bar{h} \in [0, I]$, taking conditional expectations, and using operator monotonicity of $\log$ together with $\log(1 - y) \le -y$,

$$\log \mathbb{E}[e^{-\Omega_s/\bar{h}} \mid \mathcal{D}_s] \preceq \log\Big(I - \tfrac{1-e^{-1}}{\bar{h}} \mathbb{E}[\Omega_s \mid \mathcal{D}_s]\Big) \preceq -\frac{1 - e^{-1}}{\bar{h}} \mathbb{E}[\Omega_s \mid \mathcal{D}_s] \preceq -\frac{a_*}{\bar{h}} \mathrm{diag}(W_s, W_s).$$

The matrix inside the logarithm is at least $e^{-1} I \succ 0$. Lemma 6 in dimension $2d_c$ with $\Upsilon_s = -\Omega_s/\bar{h}$, $\Theta_s = -a_* \mathrm{diag}(W_s, W_s)/\bar{h}$, and $\delta_1 = 2d_c e^{-\ell_{\mathrm{sh}}}$ shows that, outside an event of probability at most $2d_c e^{-\ell_{\mathrm{sh}}}$,

$$V_{t,e}^c - \lambda_c I \succeq \sum_{s<t} \Omega_s \succeq a_* \mathrm{diag}(S_t, S_t) - hK_* \ell_{\mathrm{sh}} I \quad \text{for all } t \ge 1. \tag{21}$$

The first inequality holds because removing the truncation indicator only adds positive semidefinite terms. Since also $V_{t,e}^c - \lambda_c I \succeq 0$, any convex combination with weight $\vartheta \in [0, 1]$ on Eq. (21) gives $V_{t,e}^c \succeq (\lambda_c - \vartheta hK_* \ell_{\mathrm{sh}})I + \vartheta a_* \mathrm{diag}(S_t, S_t)$. With $\vartheta = \lambda_c/(\lambda_c + hK_* \ell_{\mathrm{sh}})$, $\lambda_c - \vartheta hK_* \ell_{\mathrm{sh}} = \vartheta \lambda_c$, so

$$V_{t,e}^c \succeq \vartheta \lambda_c I + \vartheta a_* \mathrm{diag}(S_t, S_t) \succeq \vartheta a_* \mathsf{D}_{t,e} = c_- \mathsf{D}_{t,e},$$

using $a_* \le 1$ and $\vartheta a_* = a_*/(1 + K_* x_\delta) = c_-$.

Upper bound. For $0 \le \theta < 1/(2\sigma^2)$, integrating the conditional moment generating function of $\sqrt{2\theta}\, g\, \varepsilon_{s,e}$ against an independent standard normal g gives

$$\mathbb{E}[e^{\theta \varepsilon_{s,e}^2} \mid \mathcal{D}_s] \le (1 - 2\sigma^2\theta)^{-1/2}. \tag{22}$$

Let $W_s = \sum_i q_i p_i p_i^\top$ be an eigendecomposition with $q_i \in [0, h]$, which is $\mathcal{D}_s$-measurable. With $\theta_i = q_i/(4\sigma^2 h) \le 1/(4\sigma^2)$, Eq. (22) and $-\frac{1}{2}\log(1-y) \le y$ for $y \in [0, 1/2]$ give

$$\log \mathbb{E}\big[e^{\varepsilon_{s,e}^2 W_s/(4\sigma^2 h)} \mid \mathcal{D}_s\big] = \sum_i \log \mathbb{E}[e^{\theta_i \varepsilon_{s,e}^2} \mid \mathcal{D}_s]\, p_i p_i^\top \preceq \sum_i \frac{q_i}{2h} p_i p_i^\top = \frac{W_s}{2h}.$$

Lemma 6 in dimension d_c with $\Upsilon_s = \varepsilon_{s,e}^2 W_s/(4\sigma^2 h)$, $\Theta_s = W_s/(2h)$, and $\delta_1 = d_c e^{-\ell_{\mathrm{sh}}}$ shows that, outside an event of probability at most $d_c e^{-\ell_{\mathrm{sh}}}$,

$$\sum_{s<t} \varepsilon_{s,e}^2 W_s \preceq 2\sigma^2 S_t + 4\sigma^2 h \ell_{\mathrm{sh}} I \quad \text{for all } t \ge 1. \tag{23}$$

For $u, v \in \mathbb{R}^{d_c}$, the quadratic form of one design increment is $\omega(n_{s,e})\mathbb{I}\{e \in \mathcal{O}_s\}\big(\varphi_s^\top u + (m_s + \varepsilon_{s,e})\varphi_s^\top v\big)^2$, and

$$\big(a + (m+\varepsilon)b\big)^2 \le 2(a+mb)^2 + 2\varepsilon^2 b^2 \le 2(1+Z_0^2)(a^2+b^2) + 2\varepsilon^2 b^2, \qquad 0 \le m \le Z_0.$$

Summing and applying Eq. (23),

$$\begin{aligned} V_{t,e}^c &\preceq \lambda_c I + 2(1+Z_0^2)\mathrm{diag}(S_t, S_t) + \mathrm{diag}\big(0,\, 4\sigma^2 S_t + 8\sigma^2 h \ell_{\mathrm{sh}} I\big) \\ &\preceq (\lambda_c + 8\sigma^2 h \ell_{\mathrm{sh}}) I + \{2(1+Z_0^2) + 4\sigma^2\}\mathrm{diag}(S_t, S_t). \end{aligned}$$

Since $C_+ \lambda_c \ge 2(1+Z_0^2)\lambda_c + 8\sigma^2 h\ell_{\mathrm{sh}} \ge \lambda_c + 8\sigma^2 h \ell_{\mathrm{sh}}$ and $C_+ \ge 2(1+Z_0^2) + 4\sigma^2$, the right-hand side is at most $C_+ \mathsf{D}_{t,e}$.

Union bound. The two boundaries fail with total probability at most $3d_c e^{-\ell_{\mathrm{sh}}}$ per edge, hence at most $3d_c|E|e^{-\ell_{\mathrm{sh}}} = 3\delta/4 \le \delta$ over all edges. No full-rank condition on S_t, lower bound on the weights, or independence across rounds or edges is used. ■

C.4.4 From the design sandwich to compatibility

Proof of Proposition 1 Work on the event of Lemma 20. Inverting Eq. (20) reverses the order, and conjugating by A_t preserves it. Because both coefficient blocks of $\mathsf{D}_{t,e}$ equal $\lambda_c I + S_{t,e}$,

$$A_t \mathsf{D}_{t,e}^{-1} A_t^\top = s_{t,e} I_2, \qquad s_{t,e} = \phi_c(X_t)^\top (\lambda_c I + S_{t,e})^{-1} \phi_c(X_t) > 0,$$

where positivity uses $\phi_c(X_t) \neq 0$. Hence

$$\frac{s_{t,e}}{C_+} I_2 \preceq G_{t,e} \preceq \frac{s_{t,e}}{c_-} I_2,$$

so the eigenvalues $0 < \lambda_1 \le \lambda_2$ of $G_{t,e}$ satisfy $\lambda_2/\lambda_1 \le C_+/c_-$. For a positive definite $G = \left(\begin{smallmatrix} a & c \\ c & b \end{smallmatrix}\right)$, using $ab \le (a+b)^2/4$,

$$\rho^2 = \frac{c^2}{ab} = 1 - \frac{\lambda_1\lambda_2}{ab} \le 1 - \frac{4\lambda_1\lambda_2}{(\lambda_1+\lambda_2)^2} = \Big(\frac{\lambda_2 - \lambda_1}{\lambda_2 + \lambda_1}\Big)^2.$$

Therefore

$$1 + \rho_{t,e} \ge 1 - |\rho_{t,e}| \ge \frac{2\lambda_1}{\lambda_1 + \lambda_2} \ge \frac{2c_-}{C_+ + c_-} = \kappa_\delta$$

for all $t \ge 1$ and $e \in E$, simultaneously, with probability at least $1 - \delta$. This holds for any $\lambda_c > 0$. Under Eq. (11), $x_\delta = 1/\lambda_0$ and $\kappa_\delta = \kappa_*$, which proves Proposition 1. The same argument bounds $|\rho_{t,e}|$, but only the lower bound on $1 + \rho_{t,e}$ is needed. ■

C.5 Proof of Theorem 1

Proof of Theorem 1 Let $\delta' = \delta/5$. The choice $\lambda_c = \lambda_0 h \log(20 d_c |E|/\delta) = \lambda_0 h \log(4 d_c |E|/\delta')$ is Eq. (11) at confidence level δ'. Algorithm 1 is a non-anticipating routing rule that uses the weighted cost design, so Proposition 1 gives $\mathbb{P}(\mathcal{E}_{\mathrm{sh}}(\kappa_*)) \ge 1 - \delta'$. Proposition 3 gives an event $\mathcal{E}_0$ with $\mathbb{P}(\mathcal{E}_0) \ge 1 - 4\delta'$. On $\mathcal{E}_0 \cap \mathcal{E}_{\mathrm{sh}}(\kappa_*)$, which has probability at least $1 - 5\delta' = 1 - \delta$,

$$\mathrm{Regret}_T \le B_T^{\mathrm{hp}}(\kappa_*).$$

It remains to simplify Eq. (17). Since $d_c \geq 1$ and $\log(20d_c|E|/\delta) \geq \log 20$, the constant in Eq. (16) satisfies $A_c = (1+2Z_0^2+4\sigma^2)/\{2d_c\lambda_0 \log(20d_c|E|/\delta)\} \leq A_0 := (1+2Z_0^2+4\sigma^2)/(2\lambda_0 \log 20)$, and $A_u \leq B_\phi^2/\lambda_u$. Let

$$L = \max\Big\{\log \frac{5|E|(T+1)(1+A_0)(1+B_\phi^2/\lambda_u)}{\delta},\ \log \frac{20d_c|E|}{\delta}\Big\} \geq 1,$$

which is polylogarithmic in T, $|E|$, d_c, and $1/\delta$, and satisfies $\ell_T \leq L$ and $\sqrt{\lambda_c} \leq \sqrt{\lambda_0 h}\sqrt{L}$. Using $K(\kappa_*) \leq (1+\sqrt{2})\kappa_*^{-1/2}$ for $\kappa_* \leq 1$, $\sqrt{4d_c+2} \leq \sqrt{6}\, d_c^{1/2}$, $\sqrt{d_u+2} \leq \sqrt{3}\, d_u^{1/2}$, and $d^{1/2} \leq d$ for $d \geq 1$, the four terms of Eq. (17) are bounded by

$$\begin{aligned}
&(1+\sqrt{2})\sqrt{8\bar{v}(\underline{m})}\,(\sqrt{\lambda_0 h}B_\theta + \sqrt{6})\,\kappa_*^{-1/2} d_c \sqrt{H_G|E|T}\,L,\\
&\tfrac{4}{\log 2}(Q_{\max} + \sqrt{2}B_0\sigma)\,d_c|E|\,L^{3/2},\\
&2B_0\sqrt{c_u}\,(\sqrt{\lambda_u}B_\theta + \sqrt{3}\,\sigma)\,H_G d_u \sqrt{T}\,L,\\
&\sqrt{2}B_0\sigma\,H_G\sqrt{T}\,L^{1/2}.
\end{aligned}$$

The coefficients depend only on B_ϕ, B_θ, σ, τ_0, τ_1, $\underline{m}$, $\overline{m}$, λ_u, and λ_0. Adding the four terms gives the rate in Theorem 1. ■

C.6 The compatibility event cannot be dropped

Regularization alone does not keep $1+\rho_{t,e}$ away from zero. If $\phi_c = 1$, write $s_j = \sum_{s<t} \omega(n_{s,e})\mathbb{I}\{e \in \mathcal{O}_s\} Z_{s,e}^j$ for $j = 0, 1, 2$. Then

$$V_{t,e}^c = \begin{pmatrix} \lambda_c + s_0 & s_1 \\ s_1 & \lambda_c + s_2 \end{pmatrix}, \qquad \rho_{t,e} = -\frac{s_1}{\sqrt{(\lambda_c+s_0)(\lambda_c+s_2)}}.$$

If an edge is repeatedly observed at the same load $z_0 > 0$ with a fixed weight, then $\rho_{t,e} \to -1$: the value $a + bz_0$ is estimated accurately, while a and b are not separated. The next proposition shows that the rectangle can then fail.

Proposition 4 (Failure without compatibility) *There is an instance satisfying Assumptions 1–3 and a fixed tie-breaking rule for the routing program such that, for every $\lambda_c > 0$ and $(\tau_0, \tau_1) \neq (0,0)$, Algorithm 1 with $\eta_{t,e} = \alpha_{t,e}^c(\delta)$ has $\mathrm{Regret}_T = T/2$ for every T and every $\delta \in (0,1)$. Along its trajectory, $1 + \rho_{t,1} = \lambda_c/\{\lambda_c + (t-1)\omega(1)\}$ for all $t \geq 1$.*

Proof Take two parallel edges from o to d, one commodity with $R_t = M_t = 1$, $\phi_u = \phi_c = 1$ (so $d_u = d_c = 1$), $\gamma_e = 0$, $\varepsilon_{t,e} = 0$, $\xi_{t,e} = 0$, and $c_1(z) = c_2(z) = 1 + z$, i.e., $\beta_e = (1,1)^\top$. Assumption 1 holds with $B_\phi = 1$ and $B_\theta = \sqrt{2}$, and Assumptions 2–3 hold with $\sigma = 0$ and any (τ_0, τ_1). Fix $\lambda_c > 0$ and $(\tau_0, \tau_1) \neq (0,0)$, and write $w = \omega(1) > 0$. When the surrogate program has several minimizers, choose the one with the largest flow on edge 1. Since $\Gamma_{t,e}^c \geq 0$ and $\log(|E|/\delta) > 0$, every radius satisfies $\eta_{t,e} \geq \sqrt{\lambda_c} B_\theta = \sqrt{2\lambda_c}$. The flow estimates are $\widehat{\gamma}_{t,e} = 0$, so $\widetilde{\mu}_{t,e} = 0$.

We show by induction that $n_{t,1} = 1$ for every t. Suppose this holds for rounds $1, \ldots, t-1$, and let $N = t - 1 \geq 0$. Edge 2 has never been used, so $\widehat{\beta}_{t,2} = 0$ and $G_{t,2} = I/\lambda_c$, and its lower coefficients are $[0 - \eta_{t,2}/\sqrt{\lambda_c}]_+ = 0$. Edge 1 has N observations at load $Z = 1$ with weight w, covariate $\chi = \sqrt{w}(1,1)^\top$, and response $y^c = \sqrt{w}\, c_1(1) = 2\sqrt{w}$. Here $A_t = I_2$, so

$$V_{t,1}^c = \lambda_c I + Nw\mathbf{1}\mathbf{1}^\top, \qquad G_{t,1} = (V_{t,1}^c)^{-1} = \frac{1}{\lambda_c}\Big(I - \frac{Nw}{\lambda_c + 2Nw}\mathbf{1}\mathbf{1}^\top\Big),$$

and $\widehat{\zeta}_{t,1} = G_{t,1}\, 2Nw\mathbf{1} = \{2Nw/(\lambda_c + 2Nw)\}\mathbf{1}$. Each coordinate of $\widehat{\zeta}_{t,1}$ is below one, while $\eta_{t,1}\sqrt{(G_{t,1})_{jj}} \geq \sqrt{2(\lambda_c + Nw)/(\lambda_c + 2Nw)} > 1$. Hence both lower coefficients on edge 1 are zero as well. The surrogate objective is then identically zero, every feasible flow minimizes it, and the tie-breaking rule sends the unit demand to edge 1. This completes the induction.

The true objective with flow n on edge 1 is $L_t = n(1+n) + (1-n)(2-n) = 2n^2 - 2n + 2$, minimized at $n = 1/2$ with value $3/2$, whereas the algorithm pays 2. Hence $\Delta_t = 1/2$ and $\mathrm{Regret}_T = T/2$.

Finally, $(G_{t,1})_{12} = -Nw/\{\lambda_c(\lambda_c + 2Nw)\}$ and $(G_{t,1})_{11} = (G_{t,1})_{22} = (\lambda_c + Nw)/\{\lambda_c(\lambda_c + 2Nw)\}$, so $1 + \rho_{t,1} = 1 - Nw/(\lambda_c + Nw) = \lambda_c/(\lambda_c + Nw)$. ■

In this instance the loads never vary, so Assumption 4 fails and no horizon-independent $\kappa > 0$ satisfies $\mathcal{E}_{\mathrm{sh}}(\kappa)$. With $\kappa = \lambda_c/\{\lambda_c + (T-1)w\}$, of order $1/T$, Proposition 3 still applies, but $K(\kappa)$ grows as $\sqrt{T}$ and the bound is not sublinear, so the example is consistent with it. In particular, the tuning of Theorem 1 does not prevent this failure. A positive marginal variance of $U_{t,e}$ need not suffice either: variation fully explained by X_t leaves no residual excitation, which is why Assumption 4 is stated conditionally on $\mathcal{D}_t$.

D EXPERIMENTAL DETAILS

This appendix lists the settings of Section 5 and reports the full results of the small-network experiment.

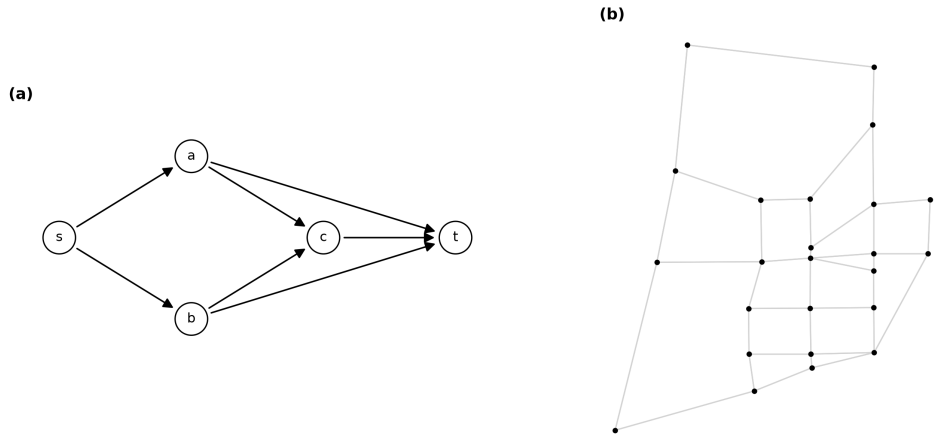


Figure 4: Networks used in Section 5. (a) Small network with 5 nodes and 7 edges; commodities $s \to t$ and $a \to t$ share the bottleneck edge $c \to t$. (b) Sioux Falls network with 24 nodes and 76 directed edges.

D.1 COMMON PROTOCOL

Algorithm. Both regressions use $\lambda_u = \lambda_c = 0.5$. The prediction of background flow is $\widetilde{\mu}_{t,e} = [\widehat{\mu}_{t,e}]_+$, and the lower corners are computed exactly as in Algorithm 1, with no bound on the cost parameters given to the algorithm. The exploration radius is $\eta_{t,e} = \eta_0\, \alpha_t^c$, where α_t^c is the self-normalized radius with $\delta = 0.1$ and $\eta_0 < 1$ is a fixed scaling constant (Table 1). The theoretical radius is calibrated for worst-case coverage and is overly conservative at these scales; rescaling it is standard practice in bandit experiments (Li et al., 2010). Naive UCB uses the same regressions, background-flow prediction, and rescaled radius. The routing program is solved by Frank–Wolfe with exact line search, as described in Appendix B, for at most 200 iterations or until the duality gap falls below 10^{-9}.

Baselines. The ε_t-greedy baseline uses the exploration probability $p_t = \min\{1, c/\sqrt{t}\}$. In an exploration round, it draws edge weights independently from $\mathrm{Unif}(0,1)$ and routes the whole demand of each commodity along the shortest path under these weights. In other rounds, it runs the same routing program with the point estimates $([\widehat{a}_{t,e}]_+, [\widehat{b}_{t,e}]_+)$. It uses the same regressions and the same background-flow prediction as RECTCOR. Setting $c = 0$ gives the certainty-equivalent greedy rule.

Naive UCB. For each edge, the minimum of $(1, z)\zeta$ over $\mathcal{C}_{t,e}^{\zeta}$ is $\ell_{t,e}(z) = \widehat{a}_{t,e} + \widehat{b}_{t,e} z - \eta_{t,e}\|(1, z)\|_{G_{t,e}}$, and the routing objective is $\sum_e g_{t,e}(n_e)$ with $g_{t,e}(n) = n\, \ell_{t,e}(n + \widetilde{\mu}_{t,e})$. Since

$\ell_{t,e}$ is concave, $g''_{t,e}(n) = 2\ell'_{t,e}(z) + n\,\ell''_{t,e}(z)$ can be negative. In each round, we check the convexity of every $g_{t,e}$ on $[0, M_t]$ on a grid of 201 points. If all are convex, a local solver (SLSQP over path flows) returns the global minimum. Otherwise, we replace each nonconvex $g_{t,e}$ by its convex envelope on its current load interval, which gives a convex relaxation and a lower bound, and branch on the edge with the largest gap between $g_{t,e}$ and its envelope at the relaxed solution, until the best feasible value is within $10^{-7}M_t$ of the lower bound. Rarely observed edges keep large confidence sets, and their $g_{t,e}$ is concave on the whole interval. The relaxation then routes no flow on them, where the envelope equals $g_{t,e}$, so they seldom require branching. Branching was needed in 0.1–5.5% of rounds. In a validation run, 100 randomly initialized local solves every 50 rounds never found a lower objective than the global solution.

Timing. Computation time is the wall-clock time of the routing decision per round, averaged over rounds and seeds. All times are measured single-threaded in Python 3.11 with NumPy and SciPy on an Intel Xeon processor at 2.8 GHz. The time of RECTCOR includes estimation and Frank–Wolfe routing; the time of naive UCB includes estimation, the convexity check, and the global solve.

Evaluation. The benchmark $f_t^\star$ is computed by the same oracle with the true parameters, and the one-round regret is the summand $\Delta_t = L_t(f_t) - L_t(f_t^\star)$ of Eq. (3). We clip Δ_t at zero to remove negative values caused by the finite oracle tolerance. Each seed draws the true parameters, and all algorithms within a seed share the same contexts, demands, background-flow residuals, and standardized cost noise. Exploration randomness of ε_t-greedy is drawn from a separate generator.

Table 1: Settings of the two experiments.

	Small network	Sioux Falls
Nodes, edges, commodities	5, 7, 2	24, 76, 10
Horizon T, seeds	4000, 20	10000, 20
Features $\phi_c = \phi_u$	$(1, x_1, x_2)$ with $x_1, x_2 \sim \mathrm{Unif}(0,1)$ i.i.d.	
Background-flow residual	truncated $\mathcal{N}(0, \sigma^2)$, $\sigma \in \{0.25, 0.5, 1\}$	$\mathcal{N}(0, 0.15^2)$ (vehicles)
Cost noise (τ_0, τ_1)	$(0.05, 0.1)$	$(0.5, 1)$
Radius scaling η_0	0.25	0.1
ε_t-greedy constant c	$\{0, 0.25, 0.5, 1, 2\}$	0.15

Background flow. In the small network, the residual $\varepsilon_{t,e}$ is drawn from $\mathcal{N}(0, \sigma^2)$ truncated symmetrically to $[-\mu_e(X_t), \mu_e(X_t)]$, independently across edges and rounds. Symmetry keeps $\mathbb{E}[\varepsilon_{t,e} \mid \mathcal{D}_t] = 0$, so $\mu_e(X_t)$ remains the conditional mean, and truncation guarantees $U_{t,e} \geq 0$ without clipping. A symmetrically truncated Gaussian is sub-Gaussian with variance proxy σ^2, and its variance is bounded below by a positive constant because $\mu_e(X_t) \geq 1.5$ on every edge. Hence the residual satisfies both the sub-Gaussian and the residual-variation assumptions. For $\sigma \leq 1$, truncation affects fewer than 0.5% of the draws, even on the edge with the smallest μ_e. In the Sioux Falls network, $U_{t,e} = \max\{\mu_e(X_t) + \varepsilon_{t,e}, 0\}$ with $\sigma = 0.15$ vehicles, while every $\mu_e(X_t)$ exceeds 290 vehicles, so the clipping never binds in practice.

D.2 SMALL NETWORK

The network has edges $s \to a$, $s \to b$, $a \to c$, $b \to c$, $c \to t$, $a \to t$, and $b \to t$. Commodity $s \to t$ has four simple paths and commodity $a \to t$ has two, so $H_G = 3$. The edge $c \to t$ lies on three of the six paths. For each seed, the coefficients are drawn once and held fixed over the horizon. On the bottleneck edge $c \to t$, each coordinate of $\beta_{e,1}$ and $\beta_{e,2}$ is drawn from $\mathrm{Unif}(0.15, 0.35)$ and each coordinate of γ_e from $\mathrm{Unif}(2.5, 4)$. On the other edges, the ranges are $(0.05, 0.25)$ for $\beta_{e,1}$, $(0.02, 0.10)$ for $\beta_{e,2}$, and $(1.5, 2.5)$ for γ_e. All coefficients are nonnegative, so a_e, b_e, and μ_e are nonnegative for every context.

In Types 1 and 2, the demands are drawn each round as $R_{t,1} \sim \mathrm{Unif}(1, 2)$ and $R_{t,2} \sim \mathrm{Unif}(0.3, 0.8)$, independently. In Type 3, they are fixed at their means $(1.5, 0.55)$. In Types 2 and 3, the context is fixed at $x = (0.5, 0.5)$.

Additional observations. As σ increases from 0.25 to 1, the mean cumulative regret of RECTCOR falls from 31.0 to 5.4 in Type 2 and from 38.4 to 20.7 in Type 1, besides the Type 3 result in the main text. The regret of naive UCB lies in 21.5–22.3 in Type 1 and 9.0–9.9 in Types 2 and 3. At $\sigma = 1$, RECTCOR attains 20.7, 5.4, and 6.4 against 21.5, 9.1, and 9.0 for naive UCB. The time ratio between the two methods is 8–16 across configurations. Because its lower bound is concave in the load, naive UCB sends whole demands along unexplored paths in early rounds; its largest single-round regret in the first 100 rounds is 5.0–6.4 on average, against 1.3–3.9 for RECTCOR. The regret of ε_t-greedy ranges from 30 to 48 and hardly depends on σ, since its exploration schedule is fixed in advance. The certainty-equivalent rule ($c = 0$) is the best method in Type 1, where context diversity alone suffices (Bastani et al., 2021), but in Types 2 and 3 it fails on some seeds, with cumulative regret up to 2092.

Table 2 reports the final cumulative regret of all methods. The ε_t-greedy variants with $c \in \{0.5, 1, 2\}$ are uniformly worse than $c = 0.25$, with mean regret between 60 and 222, and are omitted from the table.

Table 2: Cumulative regret at $T = 4000$ on the small network: mean and maximum over 20 seeds. Mean routing time per round: RECTCOR 0.30 ms, and naive UCB 2.97 ms (1.7–9.9 ms across runs).

		RECTCOR		Naive UCB		Greedy ($c = 0$)		ε_t-greedy	
Type	σ	mean	max	mean	max	mean	max	mean	max
1	0.25	38.4	51	22.3	28	9.7	81	35.8	61
1	0.5	28.3	49	21.9	27	6.6	48	32.5	47
1	1	20.7	32	21.5	32	3.2	5	31.7	47
2	0.25	31.0	240	9.7	12	93.5	1761	48.3	328
2	0.5	11.2	28	9.2	12	209.6	2092	35.2	76
2	1	5.4	14	9.1	12	6.8	74	30.9	47
3	0.25	24.4	66	9.9	12	70.2	780	36.7	74
3	0.5	10.4	28	9.2	11	102.9	2003	37.0	148
3	1	6.4	23	9.0	11	18.5	329	30.1	47

D.3 SIOUX FALLS NETWORK

Network and demand. We use the Sioux Falls network and trip table from Transportation Networks for Research Core Team (2016). We select 10 origin–destination pairs among those whose free-flow shortest path has at least three edges. The selection is greedy: it favors pairs whose paths share some edges with the pairs already selected, while also covering new edges. The selected pairs are $(10, 20)$, $(7, 12)$, $(12, 7)$, $(13, 17)$, $(17, 13)$, $(1, 19)$, $(7, 14)$, $(16, 23)$, $(9, 19)$, and $(4, 19)$, with base demands 2500, 700, 700, 500, 500, 300, 200, 500, 400, and 200. In each round, each base demand is multiplied by an independent factor drawn from $\mathrm{Unif}(4, 8)$.

Costs. Each edge has a BPR travel time $c_e^{\mathrm{BPR}}(z) = t_e^0\big(1 + b_e^{\mathrm{BPR}}(z/\mathrm{cap}_e)^{p_e}\big)$ with the free-flow time t_e^0, capacity cap_e, and parameters given in the data set. We fit an affine function $\bar{a}_e + \bar{b}_e z$ by least squares on 60 equally spaced loads in $[0, 1.5\,\mathrm{cap}_e]$ and clip both coefficients at zero. For each edge and seed, we draw $r_e^a, r_e^b, r_e^\gamma \sim \mathrm{Unif}(0.15, 0.35)$ independently and set

$$\beta_{e,1} = \bar{a}_e\big(1 - r_e^a, r_e^a, r_e^a\big), \qquad \beta_{e,2} = \bar{b}_e\big(1 - r_e^b, r_e^b, r_e^b\big).$$

Since $\mathbb{E}[\phi_c(X)] = (1, 1/2, 1/2)$, the context-averaged cost equals the fitted affine cost.

Background flow. We compute the user-equilibrium flow $\bar{n}_e$ for the full trip table of 528 origin–destination pairs under BPR costs, using 100 Frank–Wolfe iterations with step size $2/(k + 2)$. The mean background flow is $\bar{\mu}_e = 0.1\,\bar{n}_e$, and $\gamma_e = \bar{\mu}_e\big(1 - r_e^\gamma, r_e^\gamma, r_e^\gamma\big)$, so that $\mathbb{E}[\mu_e(X)] = \bar{\mu}_e$. The residual variation is thus negligible relative to the loads, and compatibility in this experiment comes mainly from the demand and context variation, which change the realized loads from round to round; Proposition 3 applies whatever supplies it (Section 4.4). The ε_t-greedy constant $c = 0.15$ was chosen from a pilot run.

Implementation. Loads in the cost regression are divided by 2500 to balance the scales of the intercept and slope features, and the precision weight is evaluated at the normalized load $n_{t,e}/2500$. The estimated slope is rescaled back before routing. The term $\sqrt{\lambda_c} B_\theta$ of the radius is dropped, so the radius uses only the self-normalized term. Each Frank–Wolfe iteration solves one shortest-path problem per commodity on the graph, so no path is enumerated.

Additional observations. A log–log fit of the cumulative regret of RECTCOR over $t \geq 200$ gives a growth rate of about $t^{0.35}$, below the $\sqrt{T}$ rate of Theorem 1. The regret of ε_t-greedy grows in jumps, each occurring when a random exploration step sends a large demand along a poor route.